



Interconnexions entre épissage alternatif et chromatine

Oriane Mauger

► To cite this version:

Oriane Mauger. Interconnexions entre épissage alternatif et chromatine. Biologie cellulaire. Université Pierre et Marie Curie - Paris VI, 2014. Français. NNT : 2014PA066093 . tel-01298427

HAL Id: tel-01298427
<https://theses.hal.science/tel-01298427>

Submitted on 6 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Pierre et Marie Curie

Ecole doctorale Interdisciplinaire pour le Vivant / Complexité du Vivant

Unité Régulation épigénétique, URA CNRS 2578, Institut Pasteur

INTERCONNEXIONS ENTRE EPISSAGE ALTERNATIF ET CHROMATINE

Par **Oriane MAUGER**

THESE DE DOCTORAT DE BIOLOGIE MOLECULAIRE

Présentée et soutenue publiquement le 4 Avril 2014

Devant un jury composé de :

Dr. Eric ALLEMAND, Chargé de Recherche _____ Co-encadrant de thèse
Pr. Yannick ANDEOL, Professeur _____ Président du jury
Dr. Didier AUBOEUF, Directeur de Recherche _____ Rapporteur
Dr. Eric BATSCHE, Chargé de Recherche _____ Directeur de thèse
Dr. Edouard BERTRAND, Directeur de Recherche _____ Examineur
Dr. Benoit MIOTTO, Chargé de Recherche _____ Rapporteur
Dr. Christian MUCHARDT, Directeur de Recherche _____ Directeur de thèse

“Success is going from failure to failure without loss of enthusiasm.”

Winston Churchill

Remerciements

RESUME

Chez l'homme, l'épissage alternatif affecte presque tous les gènes du génome permettant de générer de vastes répertoires d'ARN et de protéines. L'épissage est un processus hautement régulé qui s'effectue principalement lorsque l'ARN est en cours de synthèse sur la chromatine. Un grand nombre d'études suggèrent que la chromatine et ses marques épigénétiques influencent les décisions d'épissage au locus correspondant. A l'inverse, d'autres données laissent penser que l'épissage peut moduler les marques épigénétiques. Au cours de ma thèse, j'ai étudié différentes voies de couplage entre l'épissage et la chromatine. D'une part, j'ai exploré l'impact de la méthylation de l'ADN sur la régulation de l'épissage. J'ai montré que les enzymes qui méthylent l'ADN ont un effet global sur l'épissage d'exons enrichis en méthylation. Mes données suggèrent que les protéines qui lient la méthylation de l'ADN sont impliquées dans cette régulation. D'autre part, j'ai exploré les conséquences de l'épissage alternatif sur la régulation de la chromatine en étudiant son impact sur l'expression et l'activité de deux méthyltransferases d'histone (HMTase) : G9A et SUV39H2. Les gènes *G9A* et *SUV39H2* génèrent des transcrits alternatifs dont l'expression est régulée selon les tissus. Tous les transcrits variants codent pour des protéines. La conservation des variants d'épissage de *G9A* dans des espèces et l'absence de différences dans leur activité HMTase, nous amènent à proposer que l'épissage alternatif est associé à une fonction non liée aux histones. A l'inverse, les isoformes de SUV39H2 exhibent des activités HMTases différentes et régulent l'expression de gènes cibles différents. L'ensemble de nos résultats apporte de nouvelles connexions dans le couplage épissage - chromatine et supporte un modèle où ces derniers s'auto-influencent.

Mots clés : épissage, transcription, chromatine, méthylation de l'ADN, méthylation des histones, G9A, SUV39H2

ABSTRACT

In humans, alternative splicing affects almost all genes in the genome and generates extensive repertoires of RNAs and proteins. Splicing is a highly regulated process which occurs primarily when the RNA is being synthesized on chromatin. Many studies suggest that chromatin and epigenetic marks influence splicing choices to the corresponding locus. Conversely, other data suggest that splicing can modulate epigenetic marks. During my thesis, I studied different ways of crosstalk between splicing and chromatin. First, I investigated the effect of DNA methylation on splicing regulation. I have shown that the enzymes that methylate DNA have an overall effect on the splicing of exons with enriched methylation. My data suggest that proteins which bind to methylated DNA are involved in this regulation. On the other hand, I explored the impact of alternative splicing on chromatin regulation studying its impact on the expression and activity of both histone methyltransferases (HMTase): SUV39H2 and G9A. *G9A* and *SUV39H2* generate variants transcripts whose expression is regulated according to tissues. All variants transcripts encode proteins. Conservation of G9A splice variants in species and no differences in their HMTase activity, lead us to propose that G9A alternative splicing is associated with a non-histone function. Conversely, SUV39H2 isoforms exhibit different HMTases activities, and regulate the expression of different target genes. All our results provide new connections in chromatin - splicing coupling and support a model in which they harbor self-influence.

Key words: splicing, transcription, chromatin, DNA methylation, histone methylation, G9A, SUV39H2

Thèse préparée dans l'Unité Régulation Epigénétique, URA CNRS 2578, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris, France

TABLE DES MATIERES

Résumé.....	4
Abstract.....	4
Table des matières.....	5
Table des figures / tableaux.....	8
Abbreviations	10
Introduction.....	11
CHAPITRE 1 – L'ÉPISSAGE DES ARN.....	12
1. <i>Historique</i>	12
1.1. L'état des connaissances à la fin des années 1970.....	12
1.2. La découverte de l'épissage.....	13
1.3. Les premières avancées sur la compréhension du mécanisme d'épissage	16
1.4. La découverte de l'épissage alternatif.....	16
2. <i>Le mécanisme d'épissage</i>	17
2.1. Les différentes réactions d'épissage	17
2.1.1. Les réactions d'auto-épissage.....	17
a. L'épissage des introns du groupe I	17
b. L'épissage des introns du groupe II.....	18
2.1.2. L'épissage des ARN de transfert	20
2.1.3. L'épissage des ARN pré-messagers	21
a. Les séquences réactives intervenant lors de l'épissage.....	21
b. Les réactions de trans-estérification	21
2.2. Les spliceosomes	22
2.2.1. Le cycle du spliceosome.....	23
a. Cas des introns courts : la définition des introns	24
b. Cas des introns longs : la définition des exons	26
c. Plasticité de l'assemblage du spliceosome	28
2.2.2. La composition du spliceosome.....	28
2.2.3. Le site catalytique.....	30
2.2.4. La dynamique du spliceosome.....	30
3. <i>L'épissage alternatif</i>	31
3.1. Estimation du nombre de gènes affecté par l'épissage alternatif	31
3.2. Différentes classes d'événements	32
3.3. Les déterminants de l'épissage alternatif.....	33
3.3.1. Les facteurs <i>cis</i>	34
a. Les sites d'épissage	34
b. Les séquences régulatrices exoniques et introniques	34
3.3.2. Les facteurs <i>trans</i>	34
a. Les protéines SR.....	35
b. Les protéines hnRNP	36
c. Les autres régulateurs <i>trans</i> de l'épissage	38
d. La dynamique et la flexibilité des facteurs d'épissage.....	38
3.3.3. Le cœur du spliceosome	39
3.3.4. Les structures secondaires de l'ARNpré-m	39
3.3.5. Les modifications post-traductionnelles	39
3.4. Quand la décision d'épissage alternatif est elle prise ?.....	39
4. <i>Le couplage de l'épissage avec d'autres processus de l'ARN</i>	40
4.1. Le couplage de l'épissage avec la transcription.....	40
4.1.1. L'épissage peut être co-transcriptionnel.....	40
4.1.2. Le couplage transcription - épissage affecte l'épissage alternatif.....	42
a. Le modèle de recrutement	42
b. Le modèle cinétique	44
c.	44
4.1.3. Qu'est ce que l'épissage co-transcriptionnel ?.....	46
4.2. Le couplage de l'épissage avec d'autres processus de maturation des ARN	47
4.2.1. Le couplage de l'épissage et de la coiffe en 5' des ARNpré-m	47
4.2.2. Le couplage de l'épissage avec la maturation 3' terminale des ARNpré-m	47
4.2.3. Le couplage de l'épissage avec l'édition	48

CHAPITRE 2 – LA CHROMATINE.....	50
1. <i>La structure de la chromatine</i>	50
1.1. La compaction de la chromatine.....	50
1.2. L'unité de base : le nucléosome.....	50
2. <i>Le remodelage de la chromatine</i>	51
3. <i>Les variants d'histones</i>	52
4. <i>Les modifications épigénétiques des histones</i>	54
4.1. La diversité des modifications post-traductionnelles d'histones.....	54
4.1.1. La diversité à l'échelle du résidu d'acide aminé.....	54
4.1.2. La diversité à l'échelle du nucléosome.....	55
4.2. Les modificateurs d'histones.....	55
4.3. Les mécanismes associés aux fonctions des modifications d'histones.....	57
4.3.1. Perturbation de la structure de la chromatine.....	57
4.3.2. Régulation de la liaison de facteurs de la chromatine.....	58
4.4. Les modifications d'histones et la transcription.....	59
5. <i>Les modifications épigénétiques de l'ADN</i>	60
5.1. La méthylation de l'ADN.....	60
5.1.1. Localisation de la méthylation de l'ADN.....	61
a. A l'échelle nucléotidique.....	61
b. A l'échelle génomique.....	61
5.1.2. La méthylation de l'ADN est catalysée par les enzymes DNMT.....	62
a. DNMT1.....	62
b. DNMT2.....	63
c. La famille des DNMT3.....	63
5.1.3. La méthylation de l'ADN est impliquée dans la répression transcriptionnelle.....	64
5.1.4. Les lecteurs de la méthylation de l'ADN jouent un rôle important dans l'action répressive de la méthylation de l'ADN.....	65
a. La famille des protéines à domaine MBD.....	65
b. Les protéines Kaiso et Kaiso-like.....	66
c. La famille des protéines à domaine SRA.....	67
5.2. La déméthylation de l'ADN.....	68
6. <i>Les ARN non codants</i>	69
CHAPITRE 3 – LE COUPLAGE EPISSAGE - CHROMATINE.....	71
1. <i>Les exons et les introns sont associés à des marques chromatiniennes différentes</i>	71
1.1. Les nucléosomes sont préférentiellement localisés au niveau des exons.....	71
1.2. Les exons et les introns sont-ils enrichis en modifications épigénétiques d'histones différentes ?.....	72
1.3. Les exons et les introns sont-ils enrichis en variants d'histones différents ?.....	73
2. <i>La chromatine affecte l'épissage alternatif</i>	73
2.1. Le modèle de recrutement : la chromatine influence le recrutement de la machinerie d'épissage.....	74
2.1.1. Recrutement des facteurs d'épissage au promoteur.....	74
2.1.2. Recrutement des facteurs d'épissage à l'intérieur des gènes.....	74
2.2. Le modèle cinétique : la chromatine affecte la vitesse d'élongation transcriptionnelle.....	75
2.2.1. Les nucléosomes représentent une barrière pour l'ARN polymérase II.....	75
2.2.2. Des marques chromatiniennes et leurs facteurs affecteraient l'épissage en modulant la vitesse d'élongation de la polymérase II.....	77
a. Le complexe SWI/SNF.....	77
b. Les modifications d'histones.....	77
c. Les protéines Argonaute.....	79
2.3. Bilan de l'impact de la chromatine sur l'épissage alternatif.....	80
2.3.1. Les facteurs de la chromatine peuvent réguler l'épissage indépendamment de leurs fonctions chromatiniennes.....	80
2.3.2. Le modèle de recrutement et le modèle cinétique ne sont pas nécessairement exclusifs.....	80
2.3.3. Mais au final, quel est le modèle ?.....	81
Projet 1 : L'impact de la méthylation de l'ADN sur l'épissage alternatif.....	83
INTRODUCTION.....	84
MATERIELS ET METHODES.....	86
RESULTATS.....	90
1. <i>Méta-analyse des proportions de nucléotides dans les différents éléments d'un gène</i>	90
2. <i>Analyse des niveaux de la méthylation de l'ADN dans le gène CD44</i>	91
3. <i>Des profils d'épissage différents ne sont pas corrélés à des niveaux de méthylations d'ADN différents</i>	92
3.1. Le traitement au PMA n'affecte pas la méthylation de l'ADN à l'intérieur du gène CD44.....	92
3.2. Des cellules différentes avec différents profils d'épissage alternatif de CD44 ont des profils de méthylation similaires.....	93

4.	<i>Impact des changements de méthylation de l'ADN sur l'épissage alternatif de CD44.....</i>	94
4.1.	Impact du traitement à l'azadeoxycytidine sur l'épissage de CD44	94
4.2.	Impact de la déplétion des enzymes DNMT sur l'épissage de CD44	94
4.3.	Impact de la déplétion des protéines MBD sur l'épissage de CD44	95
5.	<i>Impact des enzymes DNMT sur l'épissage à l'échelle du génome.....</i>	96
5.1.	Impact de la déplétion de l'enzyme DNMT1 dans les cellules HeLa	96
5.2.	Impact de l'inactivation des enzymes DNMT1, DNMT3A et DNMT3B dans les cellules ES	98
5.2.1.	Analyse du transcriptome des cellules ES invalidées pour les DNMT	98
5.2.2.	Analyse du méthylome des cellules ES	99
6.	<i>Conclusions.....</i>	100
	DISCUSSION ET PERSPECTIVES	118
1.	<i>L'interprétation du couplage épissage - chromatine est complexe.....</i>	118
1.1.	Corrélation n'implique pas causalité	118
1.2.	Les facteurs chromatinien peuvent réguler l'expression des facteurs d'épissage	120
1.3.	Les facteurs chromatinien peuvent réguler les modifications post-traductionnelles de facteurs d'épissage... ..	122
1.4.	Autorégulation de l'épissage et de la chromatine	123
1.5.	Bilan	124
2.	<i>Des approches expérimentales permettant de limiter les paramètres non contrôlables doivent être développées.....</i>	125
	Project 2: Alternative splicing extends functions of G9A and SUV39H2 histone methyltransferases	127
	INTRODUCTION	128
	MATERIALS ET METHODS	130
	RESULTS	135
1.	<i>Alternatively splicing affects a significant part of G9A and SUV39H2 transcripts</i>	135
2.	<i>Alternative spliced transcripts of G9A and SUV39H2 engender variant proteins.....</i>	137
3.	<i>Alternative splicing regulates stability of G9A and SUV39H2 proteins</i>	138
4.	<i>Alternative splicing regulates sub-nuclear localization of SUV39H2 proteins.....</i>	139
5.	<i>Part of SUV39H2 HMTase does not carry enzymatic activity</i>	140
6.	<i>SUV39H2 isoforms have different target genes</i>	141
	DISCUSSION	162
	Bibliographie	166

TABLE DES FIGURES / TABLEAUX

Introduction

Figure 1. Le gène et l'ARNm hexon de l'adénovirus 2 ne sont pas colinéaires.	14
Figure 2. La réaction de trans-estérification.	18
Figure 3. L'épissage des introns des groupes I et II.	19
Figure 4. L'épissage des introns des ARN de transfert.	20
Figure 5. L'épissage des introns des ARN pré-messagers.	21
Figure 6. Le cycle du spliceosome majeur.	27
Figure 7. La composition du spliceosome majeur.	29
Figure 8. Estimation du pourcentage de gènes affectés par l'épissage alternatif chez l'homme depuis les années 1970 jusqu'à aujourd'hui.	32
Figure 9. Les classes d'événements alternatifs menant à l'expression d'ARNm différents.	33
Tableau 1. La famille des protéines SR.	35
Tableau 2. La famille des hnRNP ayant un rôle dans la régulation de l'épissage.	37
Figure 10. Formation de boucles sur les ARN naissants.	41
Figure 11. Modèle cinétique du couplage transcription - épissage.	45
Figure 12. Le nucléosome.	51
Tableau 3. Variants d'histones décrits chez l'homme. Les informations sont issues de :	53
Figure 13. Principales modifications post-traductionnelles de queues d'histones et les principales enzymes modificatrices associées.	56
Figure 14. Domaines de liaison aux histones modifiées.	58
Figure 15. Profils des modifications d'histones associées à un méta-gène transcriptionnellement actif.	60
Figure 16. Conversion d'une cytosine en 5-méthylcytosine par une ADN méthyltransférase.	62
Figure 17. La famille des ADN méthyltransférases (DNMT).	62
Figure 18. La famille des protéines à domaine MBD.	66
Figure 19. Les protéines Kaiso et Kaiso-like.	66
Figure 20. La famille des protéines à domaine SRA.	67
Figure 21. Les différentes voies de déméthylation actives de l'ADN.	69
Figure 22. Profil du positionnement du nucléosome au niveau des exons constitutifs internes.	72
Figure 23. Profil de H3K36me3 dans les exons constitutifs internes.	73
Figure 24. Les nucléosomes représentent une barrière pour l'ARN polymérase II <i>in vivo</i> et <i>in vitro</i>	76

Résultats - Projet 1

Figure 1. Biais des proportions nucléotidiques dans les différents éléments d'un méta-gène de souris.	104
Figure 2. Les cassettes exons sont moins riches en dinucléotides CpG.	105
Figure 3. Schéma explicatif de la procédure du MeDIP.	106
Figure 4. Les exons du gène CD44 sont enrichis en méthylation de l'ADN.	108
Figure 5. Le traitement au PMA n'affecte pas la méthylation de l'ADN à l'intérieur du gène CD44.	109
Figure 6. Les profils de méthylation de l'ADN à l'intérieur du gène CD44 ne sont pas corrélés aux profils d'épissage de CD44 dans les cellules MCF7, SKOV3 et SW626.	110
Figure 7. Le traitement à l'azadeoxycytidine est inefficace sur la déméthylation intragénique de CD44.	111
Figure 8. La déplétion de DNMT1 induit une diminution de l'inclusion des exons variants de CD44.	112
Figure 9. La déplétion des protéines à domaine MBD affecte l'épissage des exons variants de CD44.	113
Figure 10. Les siRNA ciblant DNMT1 ont des effets différents.	114
Figure 11. Les siRNA ciblant DNMT1 ont des effets non spécifiques.	115
Figure 12. L'inactivation des enzymes DNMT1, DNMT3A et DNMT3B affecte le transcriptome des cellules ES murines.	116
Figure 13. Les exons régulés par l'inactivation des DNMT sont enrichis en méthylation de l'ADN dans les cellules parentales.	117

Résultats Projet 2

Figure 1. Alternative splicing regulates the expression of <i>G9A</i> and <i>SUV39H2</i> transcripts.	143
Supplementary Figure S1. Alternative splicing regulates the expression of <i>G9A</i> and <i>SUV39H2</i> transcripts.	146

Figure 2. Alternatively spliced transcripts of G9A and SUV39H2 encode proteins.	147
Supplementary Figure S2. Alternative spliced transcripts of G9A and SUV39H2 encode for protein factors. ...	148
Figure 3. Transcripts resulting from alternatively spliced exon 3 of SUV39H2 encode proteins with stability and subnuclear localization differences.	149
Supplementary Figure S3. Transcripts resulting from alternative spliced exon 3 of SUV39H2 encode for proteins with stability and subnuclear localization differences.	150
Figure 4. Inclusion of exon 3 in SUV39H2 transcripts is required to encode an active histone methyltransferase.	151
Supplementary Figure S4. Inclusion of exon 3 in SUV39H2 transcripts is required to encode for an active histone methyltransferase.	152
Figure 5. Transcriptional regulation of different target genes by SUV39H2 protein isoforms.	153
Supplementary Figure S5. The protein isoforms of SUV39H2 regulate the transcription of different target genes.	154
Figure 6. Coupling between splicing, chromatin and transcription.	155
Figure S6. Lentiviral vectors map and sequences.	157
Table S1. Genes differentially expressed according to enforced expression of SUV39H3_E3S <i>versus</i> SUV39H2_E3L.	159
Table S2. Primers sequences.	160
Table S2 bis. siRNA location and sequences.	160
Table S3. Illumina sequencing run statistics.	161
Table 1. List of known chromatin factors regulated by alternative splicing.	162

ABBREVIATIONS

5mC	5-méthylcytosine
ADN	Acide DésoxyriboNucléique
AGO	ArGOnaute
ARN	Acide RiboNucléique
ARNm	ARN messenger
ARNpré-m	ARN pré-messenger
ARNnh	ARN nucléaire hétérogène
ARNr	ARN ribosomaux
ARNt	ARN de transfert
AZAdC	AZAdeoxyCytidine
ChIP	ImmunoPrécipitation de la Chromatine
CBC	Cap Binding Complex
CPSF	Cleavage and Polyadenylation Specifity Factor
CTD	Carboxy Terminal Domain
DNMT	DNA MethylTransferase
EHMT2	Euchromatic Histone-lysine N-MethylTransferase 2
EJC	Exon Junction Complex
ESE	Exonic Splicing Enhancer
ESS	Exonic Splicing Silencer
H3K9me	méthylation de la lysine 9 de l'histone H3
HMTase	Histone MethyltransferASE
hnRNP	heterogenous nuclear RiboNucleoProtein
ISE	Intronic Splicing Enhancer
ISS	Intronic Splicing Silencer
kb	kilobase
lncRNA	long non coding RNA
MBD	methyl-CpG-binding domain
MeDIP	Immunoprécipitation de l'ADN méthylé
NMD	Nonsense-Mediated mRNA Decay
pb	paire(s) de bases
PMA	Phorbol-12-Myristate-13-acetate
PTB	Polypyrimidine Tract Binding protein
siRNA	petit ARN interférent
snRNA	small nuclear RNA
snRNP	small nuclear RiboNucleoProtein
SUV39H2	SUPpressor of Variegation 3-9 Homolog 2
SWI/SNF	SWItch/Sucrose NonFermentable
TKO	Triple Knock-Out
WT	Wild Type

INTRODUCTION

CHAPITRE 1 – L'ÉPISSAGE DES ARN

L'ARN (acide ribonucléique) est une molécule universelle chez les êtres vivants. Les ARN effectuent des fonctions variées dans la cellule : messenger, guide, régulateur ou catalytique. Pour être fonctionnels, les ARN subissent de nombreux processus de maturation.

1. HISTORIQUE

1.1. L'état des connaissances à la fin des années 1970

Il existe plusieurs classes d'ARN (Darnell et al., 1986) :

- **les ARN ribosomaux (ARNr)** : synthétisés par les ARN polymérases I et III, ils représentent environ 80% des ARN totaux.
- **les ARN de transferts (ARNt)** : synthétisés par l'ARN polymérase III, ils représentent environ 15% des ARN totaux.
- **les ARN messagers (ARNm)** : synthétisés par l'ARN polymérase II, ils servent de matrice à la synthèse protéique et sont présents dans le cytoplasme. Leur taille est comprise entre 0,5 et 3 kilobases (kb).
- **les ARN nucléaires hétérogènes (ARNnh)** : nommés ainsi en raison de l'importante hétérogénéité de leur taille, entre 2 kb et 30 kb, ils sont présents dans le noyau.

A l'instar des ARNr et des ARNt, il est proposé que les ARNm puissent eux aussi être formés à partir d'ARN précurseurs nucléaires. De manière intéressante, les ARNm et les ARNnh ont une composition en bases similaire suggérant ainsi que certains ARNnh pourraient être des précurseurs d'ARNm. (Darnell et al., 1986)

Dans les années 60 et 70, deux étapes de maturation des ARN sont mis en évidence :

- **l'ajout de la queue poly A en 3' de l'ARN :**

En exposant très brièvement (2 min ou moins) des cellules avec de l'adénosine radio-marquée au tritium (^3H), il est montré que dans le noyau, une queue poly A est ajoutée en 3' de certains ARNnh de manière post-transcriptionnelle puisque aucune longue séquence répétée d'adénosines n'est présente dans le génome. De plus, en suivant la fraction des ARN marqués

au ^3H au cours du temps il est mis en évidence que les ARNm sont issus des ARNnh. (Darnell et al., 1986)

- **l'ajout de la coiffe en 5' de l'ARN :**

La découverte de la queue poly A amène à purifier plus aisément les ARNm et les ARNnh. Leur analyse a ensuite révélé qu'ils possèdent une coiffe en 5' de l'ARN : un groupement guanosine méthylé en position N7 (m^7G). (Darnell et al., 1986)

La composition similaire en nucléotides et leurs maturations communes en 5' et 3' conduisent à proposer que les ARNnh sont les précurseurs des ARNm. Pourtant, les tailles de ces deux classes d'ARN sont très différentes : les ARNnh sont environ 10 fois plus grands que les ARNm. Cette différence s'explique par un troisième processus de maturation appelé **épissage** et qui consiste à l'excision de fragments des ARNnh.

1.2. La découverte de l'épissage

Les travaux de l'équipe de Phillip A. Sharp (Berget et al., 1977) et ceux de l'équipe de Richard J. Roberts (Chow et al., 1977), publiés en 1977, ont permis de mettre en évidence le phénomène d'épissage des ARN. Philip A. Sharp et Richard J. Roberts reçoivent en 1993 le prix Nobel de physiologie ou médecine pour cette découverte.

Cette découverte se base sur des observations obtenues par microscopie électronique de R-loop (hybride d'ARN et d'ADN double brin) et d'hybridation ARN/ADN simple brin. Ces évidences amènent l'équipe de Phillip A. Sharp à proposer que le gène hexon et son ARNm ne sont pas colinéaires. Expérimentalement, ils observent que l'extrémité 5' de l'ARNm ne s'hybride pas au fragment « HindIII A » du génome de l'adénovirus 2 qui a été cartographié comme étant le gène hexon. Philip A. Sharp et ses collaborateurs ont alors hybridé l'ARNm hexon au fragment en amont du fragment « HindIII A ». Cette expérience leur a permis d'observer des hybrides ARN/ADN entrecoupés de structures en boucle d'ADN au niveau de l'extrémité 5' de l'ARNm hexon (**Figure 1A**), c'est-à-dire l'extrémité qui ne s'hybride pas au fragment « HindIII A » précédemment utilisé. Le groupe de Philip A. Sharp en a conclu que l'extrémité 5' de l'ARNm hexon est complémentaire à 3 régions non contiguës de l'ADN de l'adénovirus 2. Un modèle a ensuite été proposé dans lequel l'ARNm hexon, en plus de l'ajout de sa coiffe en 5' et de sa queue poly A en 3', subirait un autre processus de maturation impliquant l'excision de certains fragments du transcrit primaire. Philip A. Sharp

propose pour la première fois le terme *splicing* (traduction anglaise d'« épissage ») et se questionnent quant à la généralisation de ce modèle aux ARNm eucaryotes.

De façon concomitante, l'équipe de Richard J. Roberts a abouti à la même observation en utilisant une technique similaire : elle a observé par microscopie électronique des R-loop, formées par l'ADN double brin de plusieurs gènes et par les ARNm correspondants, auxquelles sont ajoutées des sondes d'ADN simple brin (fragments de restriction du génome de l'adénovirus 2). L'extrémité 5' de l'ARNm hexon qui ne s'hybride pas dans la R-loop est capable de s'hybrider à des régions non contiguës de la sonde d'ADN en formant des boucles d'ADN (**Figure 1B**). L'ensemble de ces travaux sont obtenus dans le cadre de recherche sur l'adénovirus 2, et des résultats similaires ont été obtenus avec plusieurs gènes du génome viral.

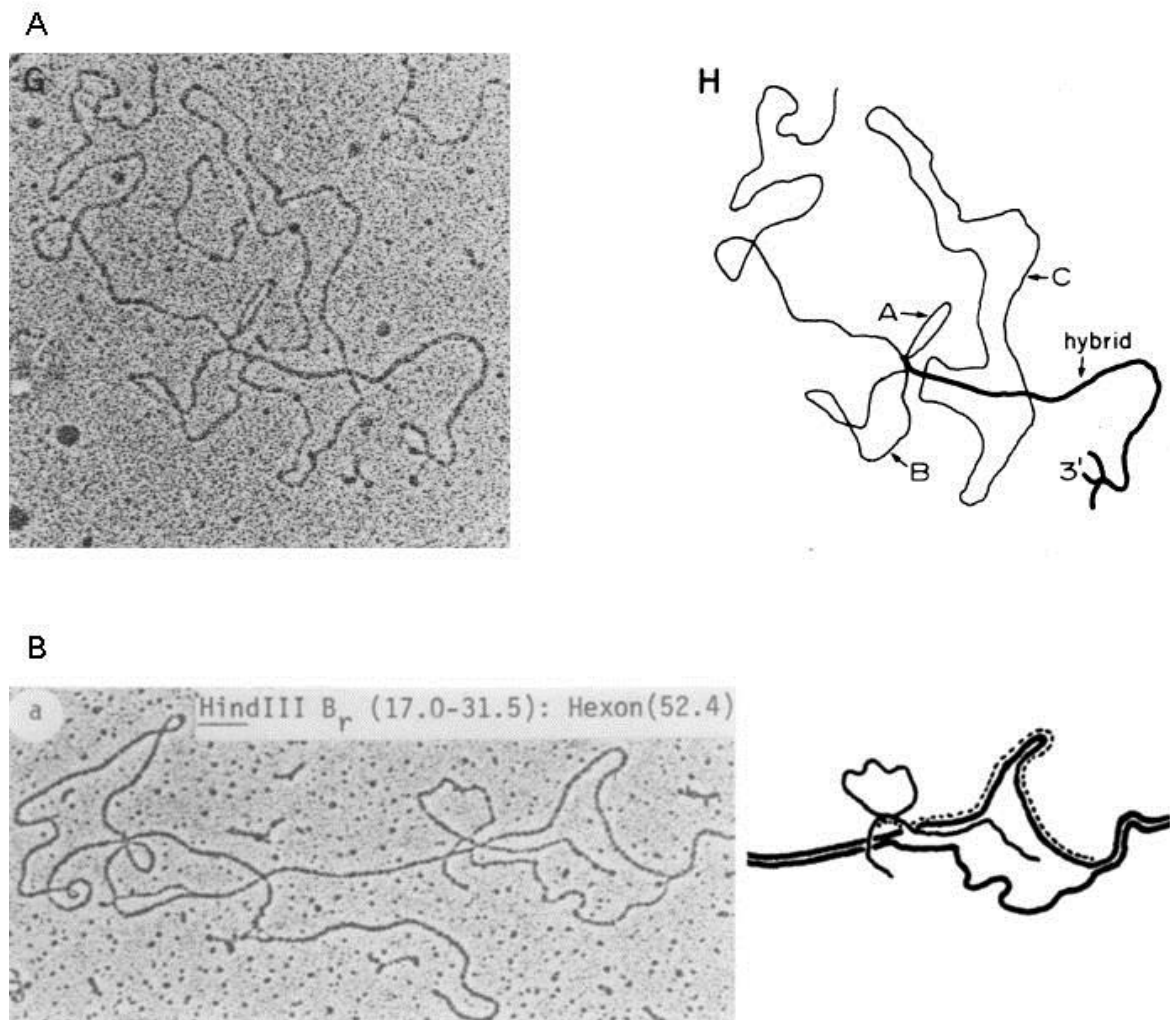


Figure 1. Le gène et l'ARNm hexon de l'adénovirus 2 ne sont pas colinéaires. **A.** Microscopie électronique d'hybrides d'ARNm hexon et de fragment d'ADN simple brin correspondant au gène hexon (gauche). Représentation schématique de l'hybride ARN/ADN (ligne épaisse) et de l'ADN non hybridé (ligne fine) (droite). (Berget et al., 1977) **B.** Microscopie électronique d'une R-loop, formée par l'ARNm hexon et par un fragment d'ADN double brin du gène hexon, hybridée à un fragment de restriction simple brin de l'ADN de l'adénovirus 2 (gauche). Représentation schématique, l'ADN est représenté par une ligne épaisse, l'ARNm par une ligne pointillée et le fragment de restriction par une ligne fine (droite). (Chow et al., 1977)

Quelques mois après ces publications, l'équipe de Pierre Chambon montre aussi la présence de « gènes rompus » (*split genes*) chez un eucaryote : le poulet. Le gène qui code pour l'ovalbumine contient 2 séquences non présentes dans l'ARNm, et au moins l'une d'elle interrompt la région codante (Breathnach et al., 1977). Pierre Chambon nomme alors ces séquences « séquences intervenantes » (*intervening sequences*). Cependant, c'est au biochimiste américain Walter Gilbert que nous devons les appellations courantes des séquences présentes ou absentes dans l'ARNm. En 1978, il dénomme les séquences exclues de l'ARNm « introns » pour la contraction d'*INTRagenic regiONs* (traduction anglaise de « régions intragéniques ») et les séquences incluses dans l'ARNm « exons » pour la contraction d'*EXpressed regiONs* (traduction anglaise de « régions exprimées ») (Gilbert, 1978).

Très rapidement, plusieurs équipes montrent d'autres exemples de « gènes rompus » dans divers organismes eucaryotes en détectant des duplex ARN/ADN, en analysant des fragments de digestion enzymatique ou par séquençage. Voici quelques exemples :

- le gène de la β -globine chez le lapin (Jeffreys and Flavell, 1977) et chez la souris (Tilghman et al., 1978a)
- les 2 gènes de la vitellogénine chez *X. laevis* (Wahli et al., 1980)
- le gène du précurseur de l'insuline chez le poulet (Perler et al., 1980) et chez l'homme (Bell et al., 1980)
- le gène du collagène de type $\alpha 2$ chez le mouton (Schafer et al., 1980)
- le gène de l'actine chez *S. cerevisiae* (Gallwitz and Sures, 1980; Jolla et al., 1980)

Malgré la caractérisation de nombreux gènes possédants des introns, beaucoup de questions restent en suspens et plusieurs d'entre elles ne sont toujours pas élucidées à ce jour (Crick, 1979) :

- Quel est le rôle des introns ? Sont-ils des éléments régulateurs ?
- Quelle est l'origine des introns ? Comment ces séquences ont-elles évolué au cours de l'évolution ?
- Comment les ARNm, composés de séquences complémentaires de régions d'ADN non contiguës, sont-ils synthétisés ? Lors de la transcription, l'ARN polymérase fait-elle des sauts au niveau des séquences introniques ? Les exons sont-ils transcrits séparément puis finalement ligués ? Est-ce qu'un transcrit primaire, colinéaire à son gène, est synthétisé avant de subir une étape d'excision des séquences introniques ?

En 1978, Tilghman *et al.* ont montré que la β -globine du lapin est d'abord transcrite sous forme d'un ARN précurseur contenant un intron. C'est la première expérience qui appuie l'hypothèse du mécanisme d'épissage. (Tilghman et al., 1978b)

1.3. Les premières avancées sur la compréhension du mécanisme d'épissage

La communauté s'est ensuite interrogée sur le mécanisme mis en jeu par l'épissage. Une des hypothèses privilégiée fut l'étude d'une complémentarité entre les extrémités 5' et 3' des introns. Ce postulat impliquait une hybridation entre ces séquences et leurs excisions par l'intermédiaire d'enzymes capables de reconnaître les extrémités des introns.

L'équipe de Pierre Chambon invalide cette hypothèse en 1978 (Breathnach et al., 1978). Les jonctions exons-introns du gène de l'ovalbumine ont été séquencées et révèlent l'absence de complémentarité entre les extrémités des introns. En revanche, cette approche a permis de découvrir un consensus aux extrémités 5' et 3' des introns, et notamment les dinucléotides GU à la frontière 5' et AG pour la jonction 3'. Ces motifs sont retrouvés dans tous les introns (ou presque...cf paragraphe 2.2. de ce chapitre).

Parce que les séquences aux abords des extrémités des introns présentent aussi un degré de conservation élevé, une hypothèse (très populaire à l'époque) fut d'imaginer que d'autres motifs puissent jouer un rôle dans le maintien des deux jonctions d'épissage. L'une des voies de recherche du moment a été d'envisager l'intervention de petits ARN nucléaires « snRNA » (*small nuclear RNA*). Il avait déjà été montré que la maturation des ARNt chez *E. coli* impliquait l'intervention d'un complexe ARN/protéine (Darnell et al., 1986). En effet, il sera montré par la suite que la réaction d'épissage est catalysée par un complexe ribonucléoprotéique. La réaction d'épissage sera présentée en détail dans le paragraphe 2.2. de ce chapitre.

1.4. La découverte de l'épissage alternatif

Dès janvier 1978, les équipes de Philip A. Sharp et Paul Berg mettent en évidence que les ARN correspondant au grand et petit antigène T du virus simien 40 (SV40) sont produits à partir d'un gène unique (Berk and Sharp, 1978; Crawford et al., 1978). Ces résultats

représentent donc la première évidence que des ARNm peuvent être le produit d'une régulation par épissage alternatif.

Le premier exemple d'épissage alternatif d'un gène eucaryote fut publié en 1982 (Amara et al., 1982) et concerne le gène de la calcitonine dont les deux ARNm distincts codent pour différentes protéines.

2. LE MECANISME D'ÉPISSAGE

2.1. Les différentes réactions d'épissage

L'épissage n'est pas une étape de maturation spécifique aux ARN pré-messagers (ARNpré-m). En effet, d'autres ARN tels que les ARNr et les ARNt sont aussi épissés. Néanmoins, en fonction du type d'ARN, il existe des différences de mécanismes catalytiques qui mènent à l'épissage.

2.1.1. Les réactions d'auto-épissage

Certains ARN possèdent intrinsèquement une activité catalytique : ils sont appelés ribozymes. Parmi les ribozymes, se trouvent les introns des groupes I et II qui sont capables de s'auto-épisser.

a. L'épissage des introns du groupe I

Les introns du groupe I sont retrouvés dans certains ARN ribosomiaux de protozoaires, dans quelques ARN mitochondriaux de champignons et dans plusieurs ARN de phages. (Nielsen and Johansen, 2009)

Ces introns possèdent de nombreuses structures secondaires (des tiges-boucles) (**Figure 3A**) et leurs structures tertiaires rapprochent des séquences éloignées dans la structure primaire, afin de former un site actif stable. La structure de ces introns est un facteur essentiel du processus de définition des sites d'épissage. (Nielsen and Johansen, 2009)

Contrairement aux structures secondaires et tertiaires, la séquence primaire de ces introns n'est pas sujette à une conservation très élevée. Néanmoins, une uridine à l'extrémité 5' de

l'exon et une guanosine à l'extrémité 3' de l'intron sont très souvent présentes. (Nielsen and Johansen, 2009)

La réaction chimique d'épissage consiste en deux réactions successives de trans-estérification (**Figure 2**). Ces dernières font intervenir une guanosine comme cofacteur et trois ions bivalents Mg^{2+} . Le 3'OH de la guanosine réalise une attaque nucléophile du site d'épissage 5', puis le 3'OH de l'uridine placée à l'extrémité 3' de l'exon, effectue à son tour une attaque nucléophile du site d'épissage 3'. (**Figure 3B**) Cette réaction ne nécessite aucune énergie extérieure, il s'agit donc d'une réaction réversible. (Saldanha et al., 1993)

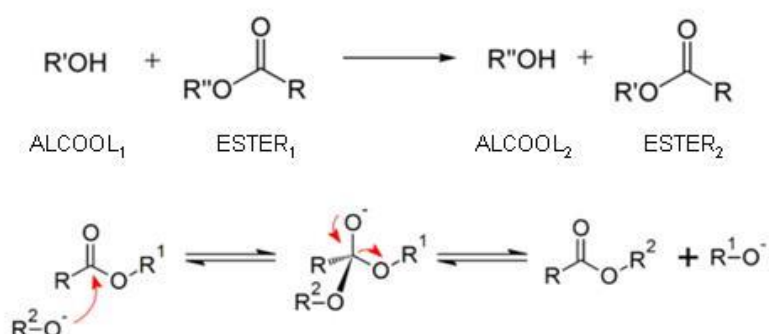


Figure 2. La réaction de trans-estérification. Un alcool de radical R' effectue réagit avec un ester de radical R'' pour former un alcool de radical R'' et un ester de radical R' (haut). Le mécanisme de la réaction est présenté (bas).

b. L'épissage des introns du groupe II

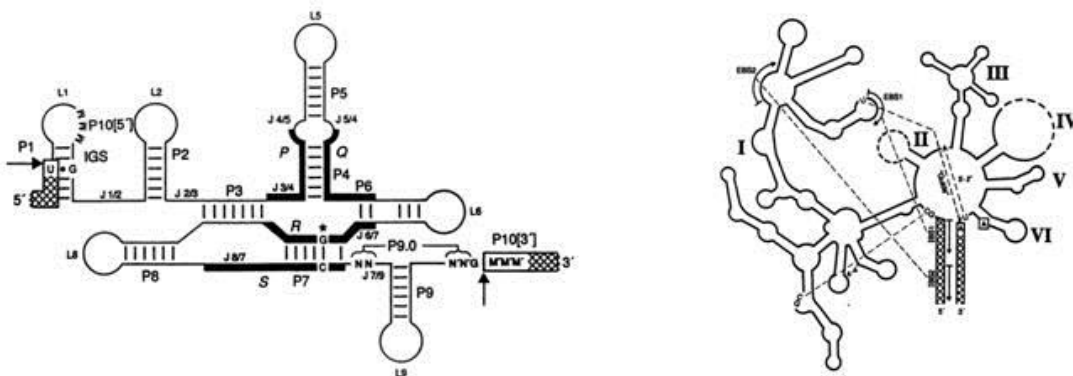
Les introns du groupe II sont présents chez les bactéries, dans certains ARN des mitochondries de champignons, des chloroplastes de plantes, des protistes et des annélides. (Lambowitz and Zimmerly, 2011)

Comme pour les introns du groupe I, la structure de ces introns joue un rôle déterminant dans la formation du site actif (**Figure 3A**). Ils interagissent avec des ions bivalents Mg^{2+} pour se lier aux séquences réactives de l'épissage : les sites d'épissage 5' et 3' ainsi que le point de branchement. (Lambowitz and Zimmerly, 2011)

De nouveau, la réaction chimique permettant l'excision des introns consiste en deux réactions successives de trans-estérification (**Figure 2**). Cette réaction est en revanche différente de celle des introns du groupe I et montre de nombreuses similitudes avec celle

effectuée pour les ARNpré-m. La première étape consiste en une attaque nucléophile du 2'OH de l'adénosine du domaine DVI (cf **Figure 3A**) sur le site d'épissage 5'. Cette réaction génère un exon 5' excisé et un intermédiaire intron/lariat/3'exon. Ensuite, le 3'OH de l'exon excisé attaque le site d'épissage 3' produisant la ligation des deux exons et libérant un intron lariat. (**Figure 3B**) De même que précédemment, cette réaction ne requière aucune énergie extérieure et par conséquent cette réaction est aussi réversible. (Lambowitz and Zimmerly, 2011)

A



B

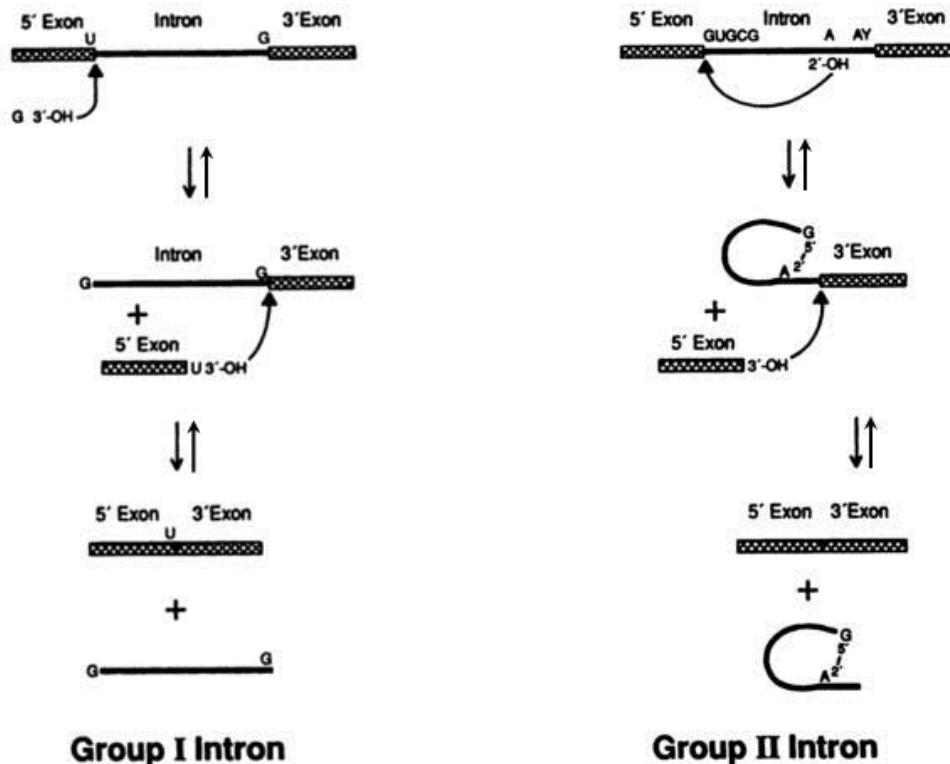


Figure 3. L'épissage des introns des groupes I et II. A. Structures secondaires des introns du groupe I (gauche) et II (droite). B. Les réactions d'épissage du groupe I (gauche) et II (droite). (Saldanha et al., 1993)

2.1.2. L'épissage des ARN de transfert

Bien que des introns soient retrouvés dans une minorité de gène d'ARNt, leur présence est toutefois montrée dans les 3 grands domaines du vivant : les Bactéries, les Archées et les Eucaryotes. Une différence existe cependant chez les Bactéries, les introns s'auto-épissent, tandis que chez les Archées et les Eucaryotes, l'épissage requièrent l'intervention de facteurs extérieurs. (Abelson et al., 1998)

Les introns des ARNt se trouvent dans la boucle contenant l'anticodon (Abelson et al., 1998). Contrairement aux autres groupes d'introns, l'épissage des ARNt ne consiste pas en une double réaction de trans-estérification.

La première étape est initiée par une endonucléase (composée chez la levure de 4 sous-unités : Sen2, Sen15, Sen34, Sen54) qui excise l'intron en laissant une demi-molécule d'ARNt 5' avec une extrémité 2'-3' phosphate cyclique et une demi molécule d'ARNt 3' avec un groupement 5'OH à son extrémité. (**Figure 4**) (Phizicky and Hopper, 2010)

La seconde étape consiste à liguer les deux demi-molécules d'ARNt. Chez la levure et la plante, cette réaction se déroule en deux temps avec une activité kinase puis ligase, suivi d'une étape finale de transfert de phosphate à l'aide d'une molécule de NAD. Chez les vertébrés et certaines archées, la ligase est capable de joindre directement les deux demi-molécules d'ARNt (**Figure 4**) (Phizicky and Hopper, 2010)

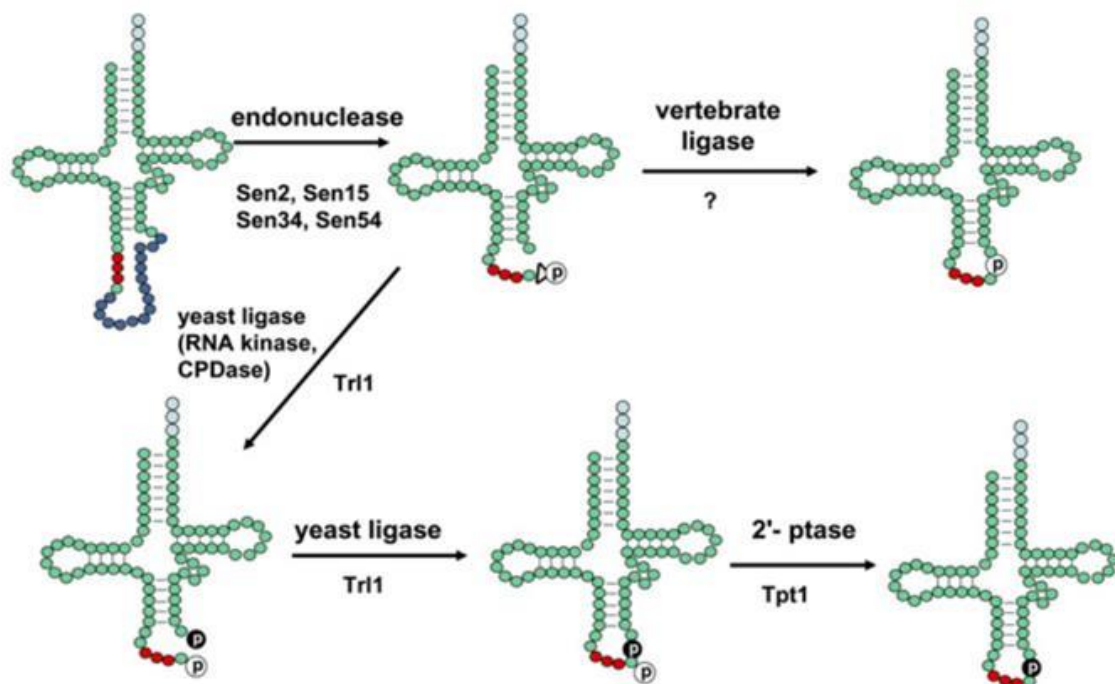


Figure 4. L'épissage des introns des ARN de transfert. L'anticodon de l'ARNt est représenté par les cercles rouges, l'intron par les cercles bleus. (Phizicky and Hopper, 2010)

2.1.3. L'épissage des ARN pré-messagers

a. Les séquences réactives intervenant lors de l'épissage

- Le site d'épissage 5'
- Le site d'épissage 3'
- Le point de branchement : il est typiquement situé entre 18 et 40 nucléotides en amont du site d'épissage 3' et d'un enrichissement de polypyrimidines.

Les séquences consensus des sites d'épissage et celle du point de branchement ont un niveau de conservation très élevé chez *Saccharomyces cerevisiae* tandis qu'elles sont beaucoup plus dégénérées chez les métazoaires (Will and Lührmann, 2011). (**Figure 5A**)

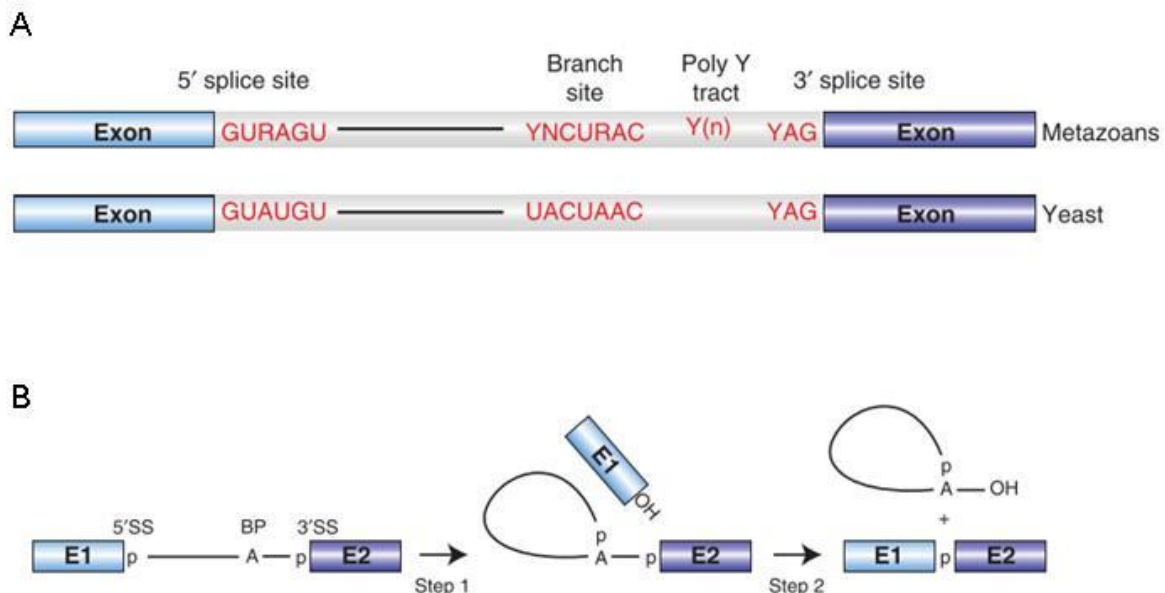


Figure 5. L'épissage des introns des ARN pré-messagers. **A.** Séquences conservées trouvées aux sites d'épissage 5' et 3' et au site de branchement des ARN pré-messagers chez la levure (*S. cerevisiae*) et chez les métazoaires. Y = pyrimidine et R = purine. L'étendue de polypyrimidine est indiquée par Y(n). **B.** Représentation schématique de l'épissage à deux étapes des ARN pré-m. Les rectangles représentent les exons (E1, E2) et les lignes fines représentent les introns. L'adénosine du site de branchement est indiquée par la lettre A. Les groupes phosphates aux sites d'épissage 5' et 3' sont indiqués par la lettre p. (Will and Lührmann, 2011)

b. Les réactions de trans-estérification

La réaction d'épissage consiste en deux trans-estérifications successives (**Figure 2**). Comme pour les introns du groupe II, la première trans-estérification s'effectue entre le groupement 2'OH de l'adénosine du point de branchement et le site d'épissage 5'. Elle conduit à l'excision de l'exon 5' et à la formation d'une structure en lasso : un intron-

lariat/exon 3'. La deuxième trans-estérification met en jeu le site d'épissage 3' et le groupement 3'OH de l'exon 5' afin de liguer les deux exons et exciser l'intron-lariat (Will and Lührmann, 2011). **(Figure 5B)**

Les introns de groupe II sont soupçonnés d'être les ancêtres des introns des ARNpré-m en raison :

- de la réaction chimique d'épissage identique,
- de la similarité des séquences mises en jeu dans des liaisons ARN-ARN,
- de la similarité structurale des domaines majeurs des introns de groupe II et des snRNA du spliceosome.

(Lambowitz and Zimmerly, 2011)

En revanche, les sites d'épissage 5' et 3' ainsi que le point de branchement des ARNpré-m contiennent de très courtes séquences consensus. De ce fait, les introns des ARNpré-m ne contiennent pas l'information suffisante pour adopter une structure secondaire et tertiaire requise à la formation du site catalytique et le rapprochement des séquences réactives. La réaction d'épissage des ARNpré-m fait donc intervenir une multitude de facteurs agissant en *trans* afin de former le **spliceosome**. Cette différence génère intrinsèquement de nombreuses possibilités de régulation. Ce mode de fonctionnement est étroitement relié au fait que l'épissage alternatif des ARNpré-m est un processus générant de la diversité dans l'expression d'un gène.

2.2. Les spliceosomes

Le spliceosome est un complexe ribonucléoprotéique de plusieurs mégadaltons qui catalyse la réaction d'épissage des ARNpré-m. Cette machinerie enzymatique macromoléculaire est la plus complexe de la cellule connue à ce jour.

Il existe deux types de spliceosome : le majeur et le mineur.

- **le spliceosome majeur** est responsable de l'excision de presque l'intégralité des introns des ARNpré-m. Il est constitué de 5 snRNP (*small nuclear RiboNucleoProtein*) U1, U2, U4/U6 et U5. Chaque snRNP comporte un snRNA (*small nuclear RNA*) et de nombreuses protéines associées. Le spliceosome est aussi

composé de nombreux autres facteurs non-associés aux snRNP. Les introns épissés par ce spliceosome sont dits de type U2.

- **le spliceosome mineur** catalyse environ 0,5% des introns (Turunen et al., 2013). Il est composé de snRNP distinctes de celles du spliceosome majeur. Elles ont toutefois des rôles analogues : U11/U12 (analogue de U1 et U2), U4atac/U6atac (analogue de U4/U6). La snRNP U5 est commune avec le spliceosome majeur. Les introns épissés par ce spliceosome sont dits de type U12.

Les introns de type U2 sont définis par les dinucléotides GU et AG aux sites d'épissage 5' et 3' respectivement. Les introns de type U12 sont quant à eux définis par les dinucléotides variables : majoritairement GU-AG et AT-AC mais d'autres combinaisons ont aussi été caractérisées. Enfin, les introns de type U12 ne possèdent pas d'enrichissement en polypyrimidines au site d'épissage 3'. De plus, la distance entre le point de branchement et le site d'épissage 3' est majoritairement comprise entre 11 et 13 nucléotides, et semble jouer un rôle important pour la reconnaissance des introns de type U12. (Turunen et al., 2013)

La signification biologique du spliceosome mineur reste une énigme.

Les processus d'épissage mis en jeu par ces deux types de spliceosomes restent néanmoins similaires, et par conséquent nous traiterons dans ce manuscrit uniquement le mécanisme d'épissage catalysé par le spliceosome majeur.

2.2.1. Le cycle du spliceosome

Le spliceosome n'est pas une enzyme existant sous la forme d'un macro-complexe pré-assemblé et catalytiquement actif sur l'ARN_{pré-m}. Il requiert un assemblage ordonné qui met en jeu une multitude d'interactions entre l'ARN_{pré-m}, les 5 snRNP et des facteurs d'épissage. Ces interactions sont établies et remaniées à l'aide de nombreux réarrangements structuraux dont les transitions et le dynamisme marquent la progression entre les différents complexes d'assemblage.

Six complexes spliceosomaux de compositions et de conformations différentes ont été caractérisés. Ils sont associés à des stades différents de l'assemblage de l'enzyme de l'épissage. Il est néanmoins admis que les réarrangements structuraux du spliceosome sont beaucoup plus nombreux, suggérant qu'il existe d'autres complexes intermédiaires. Leurs descriptions sont mal caractérisées en raison de leurs états transitoires (**Figure 6A**).

a. Cas des introns courts : la définition des introns

Dans le cas des introns courts ($\sim <250$ nt), l'assemblage du spliceosome se fait initialement autour de l'intron (Fox-Walsh et al., 2005).

Complexe E : le complexe d'engagement (*commitment complex*)

L'assemblage du spliceosome est initié par la formation du complexe E. Ce complexe joue un rôle crucial dans l'initiation de la reconnaissance des sites d'épissage 5' et 3'.

Il consiste en la liaison de la snRNP U1 au site d'épissage 5' par des appariements de bases du snRNA U1 et du site d'épissage 5'. Cette interaction est stabilisée par les protéines de la snRNP U1 ainsi que des facteurs d'épissage de la famille des protéines SR (riche en sérine/arginine) (Shepard and Hertel, 2009).

De façon concomitante, ce complexe met en jeu d'autres facteurs: la protéine SF1/BBP qui se lie au point de branchement, et l'hétérodimère U2AF65/U2AF35 qui interagit avec l'étendue de polypyrimidines et le dinucléotide AG du site d'épissage 3' (**Figure 6B**). U2AF65 peut être substitué par PUF60 (Hastings et al., 2007) et U2AF35 par U2AF26 (Shepard et al., 2002). Le complexe E est aussi décrit pour contenir la snRNP U2, mais l'interaction de cette snRNP n'implique pas d'appariement de son snRNA et uniquement des contacts protéines-protéines (Das et al., 2000). (Wahl et al., 2009; Will and Lührmann, 2011)

In fine, cette étape a pour but de rapprocher les sites d'épissage 5' et 3'.

La formation du complexe E ne nécessite pas l'hydrolyse d'ATP en opposition avec les autres étapes d'assemblage du spliceosome qui sont dépendantes de l'ATP, notamment au travers de nombreuses ARN hélicases de type DExD/H (Wahl et al., 2009; Will and Lührmann, 2011)

Complexe A : le pré-spliceosome

La transition au complexe A requière de l'énergie qui est issue de l'hydrolyse de l'ATP. Elle est principalement consommée par des hélicases et sert à catalyser l'ouverture de molécules d'ARN double brin ou déplacer des protéines liées à l'ARN. De cette manière, une région de l'ARN peut s'engager dans un nouvel appariement de bases ou dans de nouvelles interactions protéiques.

Ce complexe met en jeu les ARN hélicases Prp5 et Sub2/UAP56, la snRNP U2 *via* l'appariement de son snRNA, afin d'établir le point de branchement. De nouveau, cette interaction est stabilisée par l'intervention de protéines associées à la snRNP U2 : les

complexes SF3a/SF3b et l'hétérodimère U2AF. L'association de la snRNP U2 conduit au déplacement de SF1/BBP du point de branchement lequel lie SF3b14a/p14, marquant ainsi l'adénosine du point de branchement. **(Figure 6B)** (Wahl et al., 2009; Will and Lührmann, 2011)

Complexe B : le spliceosome pré-catalytique

Il met en jeu les snRNP U4/U6 et U5 qui sont recrutées sous forme d'une tri snRNP pré-assemblée. Le snRNA U5 établit des contacts avec des nucléotides des exons 5' et 3'. L'extrémité 3' du snRNA U6 (toujours sous forme de la di-snRNP U4/U6) s'apparie avec l'extrémité 5' du snRNA U2. Par ailleurs, l'ARN hélicase Prp28 est impliquée dans le retrait de l'extrémité 5' du snRNA U1 au site d'épissage 5' au profit du snRNA U6. Cet événement initie l'activation du spliceosome. Par ailleurs, bien que la snRNP U1 ne soit plus liée par des appariements de bases, cette snRNP est toujours présente dans le complexe B. **(Figure 6C)**

A cette étape, les 5 snRNPs du spliceosome sont présentes mais l'enzyme n'est pas catalytiquement active car elle nécessite des changements importants de sa conformation et de sa composition. Notamment, le snRNA U6 contient des éléments essentiels du site catalytique qui sont inactifs car ils sont encore appariés au snRNA U4. (Wahl et al., 2009; Will and Lührmann, 2011)

Complexe B^{act} : le spliceosome activé

La formation du complexe B^{act} correspond à l'activation du spliceosome.

Le remaniement des interactions ARN/ARN et ARN/protéines mène à la déstabilisation des snRNP U1 et U4. Les interactions entre les snRNA U4 et U6 sont rompues et le snRNA U6 s'apparie aux nucléotides du site d'épissage 5' *via* un motif conservé ACAGAG **(Figure 6C)**. D'autres duplexes U2/U6 sont formés. Successivement, l'hélicase Brr2 est requise pour la dissociation du di-snRNA U4/U6 et permet l'appariement entre les snRNA U6 et U2. (Wahl et al., 2009; Will and Lührmann, 2011)

Complexe B* : le spliceosome catalytiquement activé

L'ARN hélicase Prp2 est un déterminant de l'activation catalytique du spliceosome, qui conduit à générer le complexe B*. Le spliceosome, dans cette conformation est alors actif pour catalyser la première trans-estérification et transiter vers le complexe C. (Wahl et al., 2009; Will and Lührmann, 2011)

Complexe C : le spliceosome catalytique

Précédant la deuxième réaction de trans-estérification, le spliceosome subit de nouveau de nombreux réarrangements protéiques, impliquant notamment les ARN hélicases Prp16 et Prp22. (Wahl et al., 2009; Will and Lührmann, 2011)

Dissociation du spliceosome :

L'action de plusieurs ARN hélicases (Prp22, Prp43, Brr2, Snu114) après la deuxième trans-estérification permet de dissocier le spliceosome et laisse l'ARNm sous forme de RNP, c'est-à-dire associé à de nombreuses protéines. Les snRNP U2, U5 et U6 sont libérées et recyclées pour d'autres réactions d'épissage. L'intron lariat est aussi libéré. (Wahl et al., 2009; Will and Lührmann, 2011)

b. Cas des introns longs : la définition des exons

Chez les eucaryotes supérieurs, les introns ont une taille très variable et peuvent atteindre des milliers de nucléotides. De ce fait, la reconnaissance des introns représente un défi majeur pour le spliceosome. De plus, ces introns arborent de nombreux pseudo sites d'épissage ce qui complexifie la reconnaissance précise des jonctions intron-exon. Cette variation en taille est beaucoup moins présente pour les exons. Ils sont de taille relativement stable, avec une longueur moyenne de 120 nucléotides. Pour surmonter les contraintes de l'assemblage du spliceosome sur les introns longs, il a été proposé qu'un mécanisme de définition des exons soit aussi mis en place dans les premières étapes d'assemblage du spliceosome.

Le modèle expliquant la définition des exons implique une liaison de la snRNP U1 au site d'épissage 5' de l'exon, une interaction de U2AF65, la snRNP U2 au site d'épissage en amont de l'exon, et des séquences régulatrices exoniques reconnues par des facteurs d'épissage de la famille des protéines SR (cf. paragraphe 3.3.2.a. de ce chapitre). Ensemble, ces facteurs établissent un réseau d'interactions protéine-protéine de part et d'autre de l'exon qui favorisent leurs stabilisations et définissent les frontières de l'exon (**Figure 6D**) (Wahl et al., 2009; Will and Lührmann, 2011). Le complexe de définition de l'exon est ensuite converti en un complexe de définition de l'intron (De Conti et al., 2013).

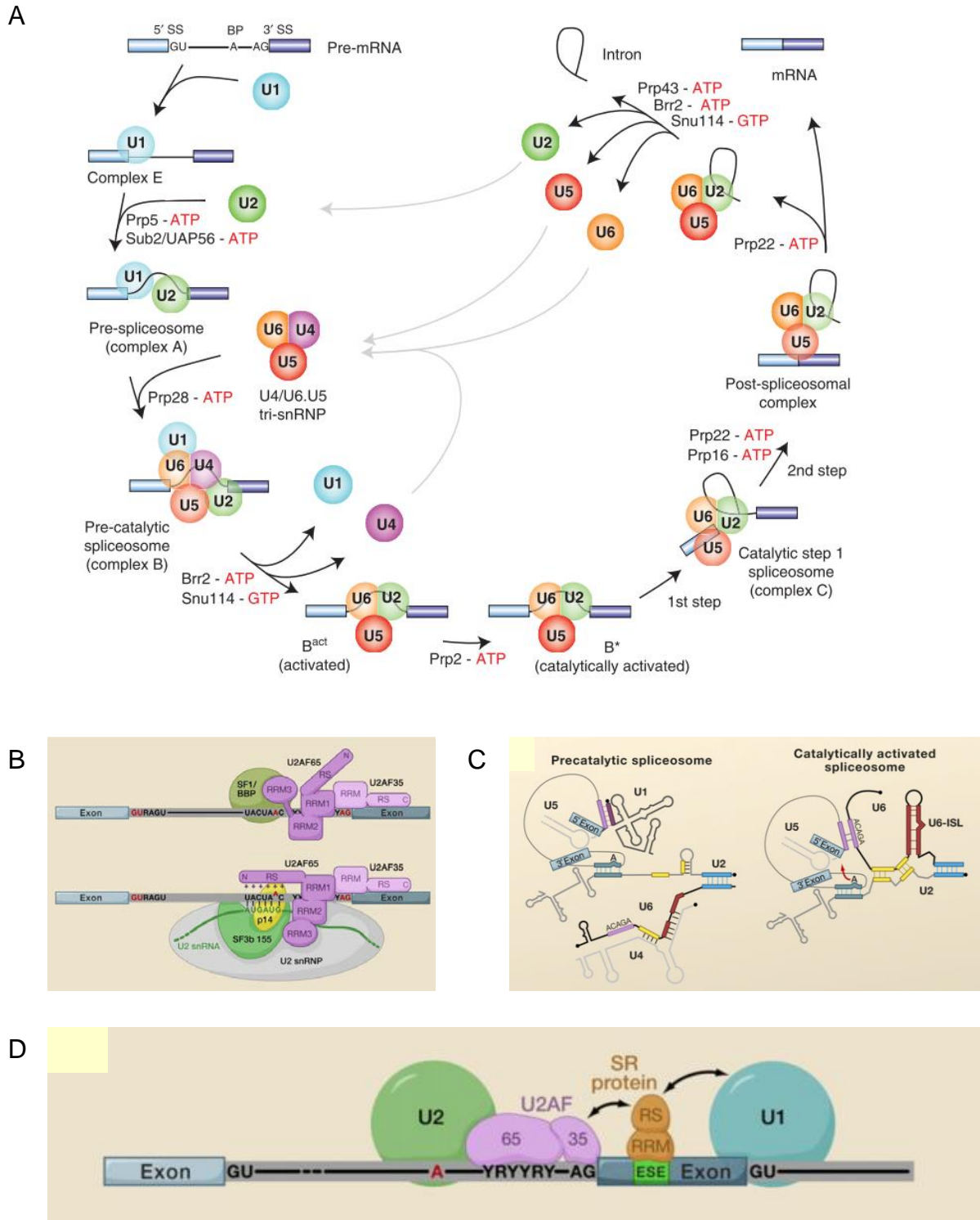


Figure 6. Le cycle du spliceosome majeur. **A.** Voie canonique d'assemblage et de dissociation du spliceosome autour de l'intron. Les exons et les introns sont représentés par des rectangles et des lignes respectivement. Les snRNP sont représentées par des cercles. (Will and Lührmann, 2011) **B.** Les interactions moléculaires au site de branchement et au site d'épissage 3' des complexes spliceosomaux E (haut) et A (bas). Les exons et les introns sont représentés par des rectangles bleus et gris respectivement. (Wahl et al., 2009) **C.** Le réseau d'interactions ARN-ARN dans le complexe pré-catalytique (B) (gauche) et dans le complexe catalytiquement actif (B*) (droite). (Wahl et al., 2009) **D.** Modèle d'interactions lors de l'assemblage précoce du spliceosome dans le cas de la définition des exons. Les exons et les introns sont représentés par des rectangles bleus et gris respectivement. (Wahl et al., 2009)

c. Plasticité de l'assemblage du spliceosome

L'assemblage du spliceosome est caractérisé par une certaine plasticité, dont la définition des exons ou des introns en sont le commencement. D'autres voies alternatives ont été suggérées, avec par exemple l'existence d'un complexe pré-assemblé contenant les 5 snRNPs, cependant ces autres possibilités sont encore sujettes aux polémiques et ne seront pas développées dans ce manuscrit (Stevens et al., 2002).

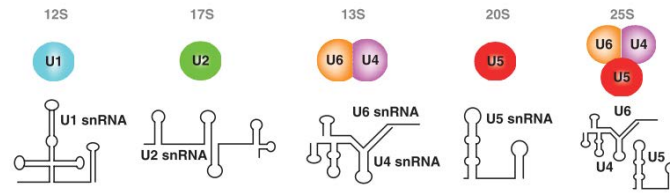
2.2.2. La composition du spliceosome

Le spliceosome est constitué de snRNP (*small nuclear RiboNucleoProtein*) U1, U2, U4/U6 et U5, et de protéines de l'épissage dites régulatrices. Chaque snRNP est formée d'un cœur composé d'un snRNA (ou deux pour la particule U4/U6), et de 7 protéines Sm communes à U1, U2, U4, U5 (B/B', D3, D2, D1, E, F et G) et LSm pour U6. Un nombre variable de protéines spécifiques sont aussi implémentés au cœur de chacune des snRNP. Le spliceosome humain contient environ 45 protéines distinctes associées aux snRNP ce qui contribue à environ 2,7MDa de sa masse moléculaire. La structure secondaire la plus probable de chaque snRNA est représentée en **Figure 7A**.

De plus, le spliceosome comprend environ 170 protéines non snRNP. Ces protéines ne sont pas toutes présentes dans le spliceosome en même temps et leurs recrutements sont dépendants de leurs rôles dans l'assemblage du spliceosome. Ainsi, chaque complexe spliceosomal contient environ 125 protéines. La régulation des échanges entre chaque état du spliceosome a un rôle fondamental dans la plasticité du spliceosome et représente un élément déterminant pour la réaction d'épissage. (Wahl et al., 2009; Will and Lührmann, 2011)

La complexité structurelle du spliceosome et son assemblage sont un défi d'organisation et de cinétique. Par conséquent de nombreuses protéines du spliceosome sont recrutées sous la forme de pré-complexes préalablement assemblés, avant leur incorporation à l'ARNpré-m et la machinerie d'épissage.

A



B

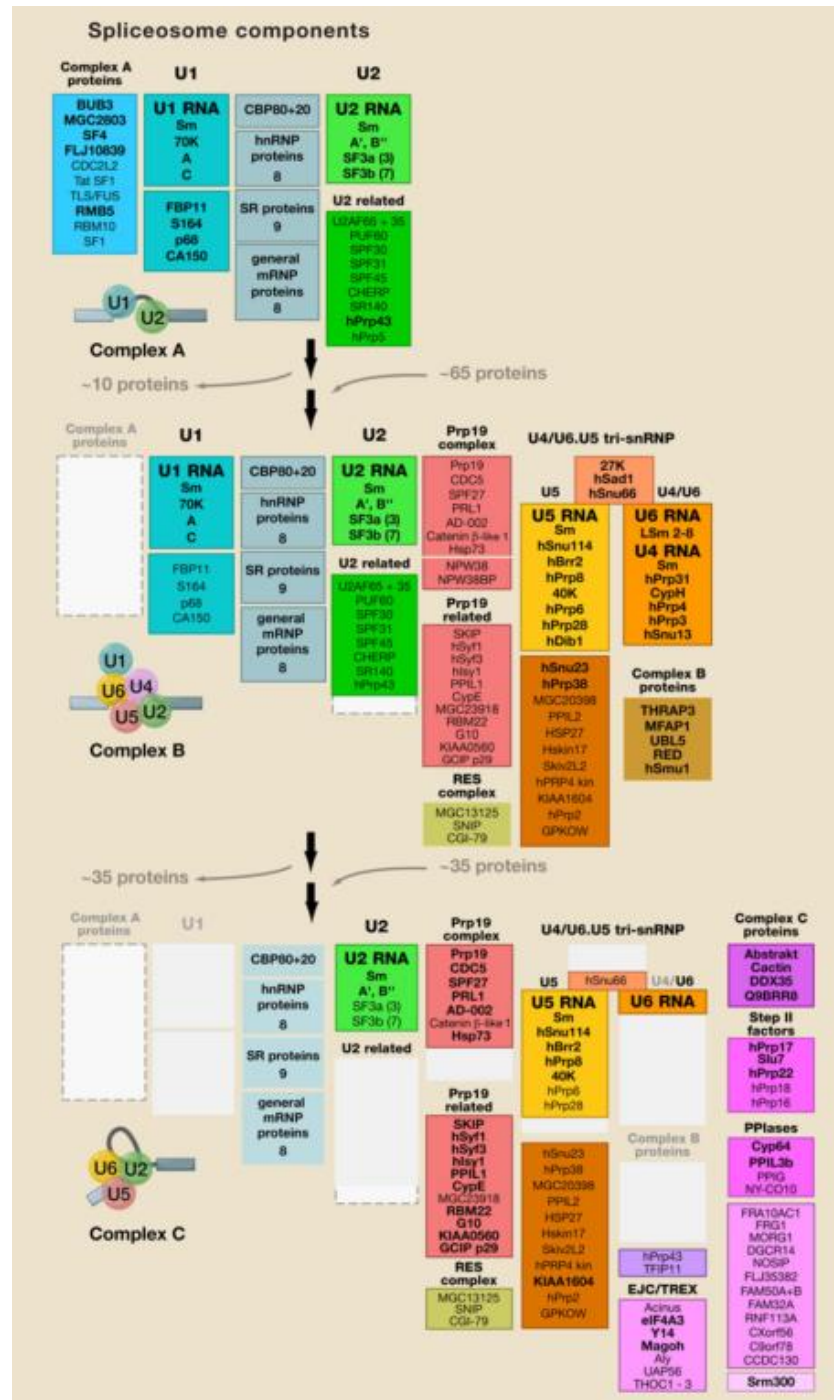


Figure 7. La composition du spliceosome majeur. A. Structures secondaires les plus probables des différents snRNA du spliceosome majeur humain. (Will and Lührmann, 2011) B. La dynamique de la composition protéines du spliceosome au cours de son cycle. (Wahl et al., 2009)

2.2.3. Le site catalytique

Comme évoqué précédemment, le site actif est initialement séquestré dans une conformation inactive (Complexes A, B, B^{act}). Ces états sont proposés pour apporter un avantage cinétique afin de laisser le temps nécessaire pour l'assemblage adéquat du spliceosome.

Par définition, le site actif d'une enzyme représente la portion où les molécules substrats se lient et subissent la réaction chimique les transformant. Dans le contexte du spliceosome, il est extrêmement difficile d'établir une distinction entre les composants structuraux et les composants actifs, c'est-à-dire ceux directement impliqués dans la double réaction de trans-estérification de l'épissage. Récemment, les travaux de S.M. Fica, N. Tuttle et de leurs collaborateurs ont montré que le snRNA U6 n'est pas uniquement un élément structural du spliceosome mais aussi un acteur majeur du cœur enzymatique en recrutant deux ions bivalents Mg^{2+} (Fica et al., 2013).

Bien que les snRNA soient depuis longtemps proposés comme les catalyseurs directs de la réaction d'épissage (par homologie avec les introns de groupe II), l'importance des facteurs protéiques ne peut toutefois pas être complètement négligée ou exclue. Plusieurs études, en particulier celles à propos de Prp8 (Grainger and Beggs, 2005), confèrent aux protéines du spliceosome un rôle dans les structures nécessaires à la catalyse de l'épissage et ces protéines sont par conséquent proposées pour avoir aussi un rôle déterminant dans la double réaction de trans-estérification.

2.2.4. La dynamique du spliceosome

Le spliceosome est une machinerie hautement dynamique caractérisée par 3 principes majeurs :

- les groupes réactifs sont reconnus de multiples fois par des ARN ou des protéines assurant la précision et la fidélité de la réaction d'épissage
- beaucoup d'interactions fonctionnellement importantes sont d'affinité individuelle faible dans un contexte binaire mais sont fortement stabilisées dans le contexte du spliceosome. Ce principe permet de la flexibilité, une propriété cruciale pour la régulation de cette machinerie (épissage alternatif)

- les réarrangements de la RNP pendant l'assemblage du spliceosome impliquent en général un échange entre un ou plusieurs partenaires à de nouveaux partenaires : i) entre ARN-ARN : snRNA-snRNA et snRNA-ARNpré-m ; ii) entre protéines-protéines.

Au moment de l'assemblage du spliceosome, des compétitions peuvent s'établir entre la sélection de plusieurs sites d'épissage 5' ou 3' se qui se concrétise par la production d'ARNm différents. Ces différents épissages sont appelés **épissages alternatifs** et sont proposés pour réguler fortement les produits de l'expression des gènes. Cette capacité de régulation est directement reliée à la dynamique des interactions de la machinerie d'épissage.

3. L'EPISSAGE ALTERNATIF

3.1. Estimation du nombre de gènes affecté par l'épissage alternatif

Considéré au départ comme une exception, il est aujourd'hui établi que l'épissage alternatif affecte la plupart des gènes chez les eucaryotes supérieurs.

Chez l'homme, les premiers exemples de gènes épissés alternativement ont été découverts au début des années 1980 et leur nombre n'a fait que croître depuis (Leff and Rosenfeld, 1986). Au début des années 1990, il était estimé que cette régulation affecte environ 5% des gènes humains (Sharp, 1994). Au cours des années 2000, le développement de bases de données d'*ESTs* (*expressed sequence tag*) et d'algorithmes d'analyses ont conduit à caractériser plus de 35% des gènes comme modulés par l'épissage alternatif (Brett et al., 2000; Mironov et al., 1999). Enfin, le développement récent de technologies d'analyses du génome à grande échelle révèle le caractère quasi-universel de l'épissage alternatif dans la régulation des gènes humains. En 2003, l'analyse du transcriptome de 52 échantillons (tissus et types cellulaires) sur des puces à jonctions d'exons, Johnson et ses collègues ont estimé qu'environ 74% des gènes humains sont épissés alternativement (Johnson et al., 2003). En 2008, les approches de séquençage à haut débit, menées par les équipes de C.B. Burge et B.J. Blencowe, ont conduit à estimer que 95% des gènes multi-exoniques ont une étape de maturation par épissage alternatif (Pan et al., 2008; Wang et al., 2008). **(Figure 8)**

Ainsi, il n'est pas impossible que l'amélioration des technologies de séquençage, conduisent à proposer que presque 100% des gènes sont affectés par l'épissage alternatif

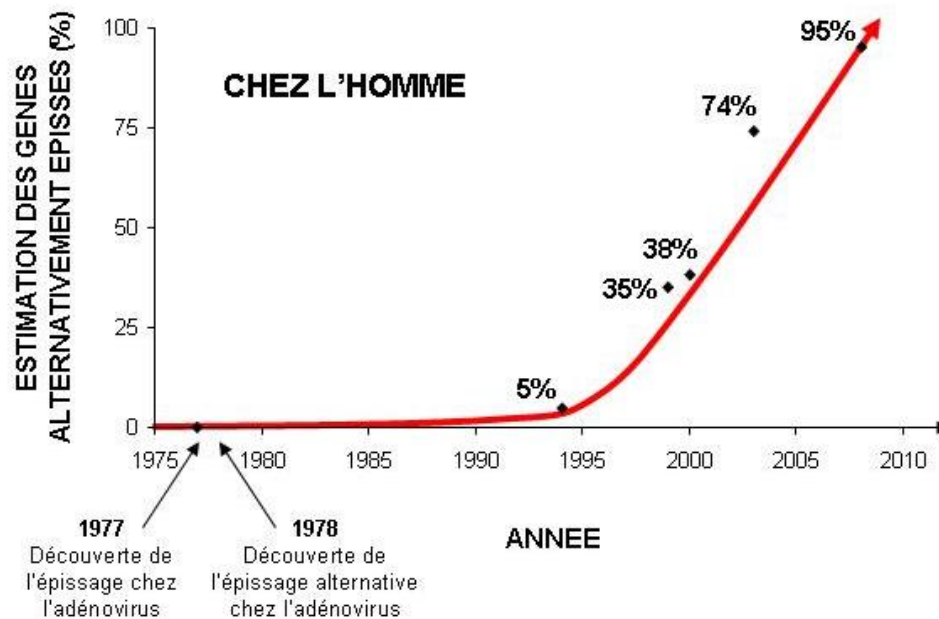


Figure 8. Estimation du pourcentage de gènes affectés par l'épissage alternatif chez l'homme depuis les années 1970 jusqu'à aujourd'hui.

3.2. Différentes classes d'événements

L'épissage alternatif implique l'utilisation différentielle de sites d'épissage. Les types d'événements d'épissage alternatifs sont variés : i) exon cassette, ii) exons mutuellement exclusifs, iii) site d'épissage 5', iv) site d'épissage 3', v) site d'épissage à spécificité duelle, vi) rétention d'intron. La diversité des ARNm générés provient aussi de l'utilisation de promoteurs ou sites de terminaison alternatifs qui sont tous deux souvent liés à des décisions d'épissage alternatives. (Figure 9)

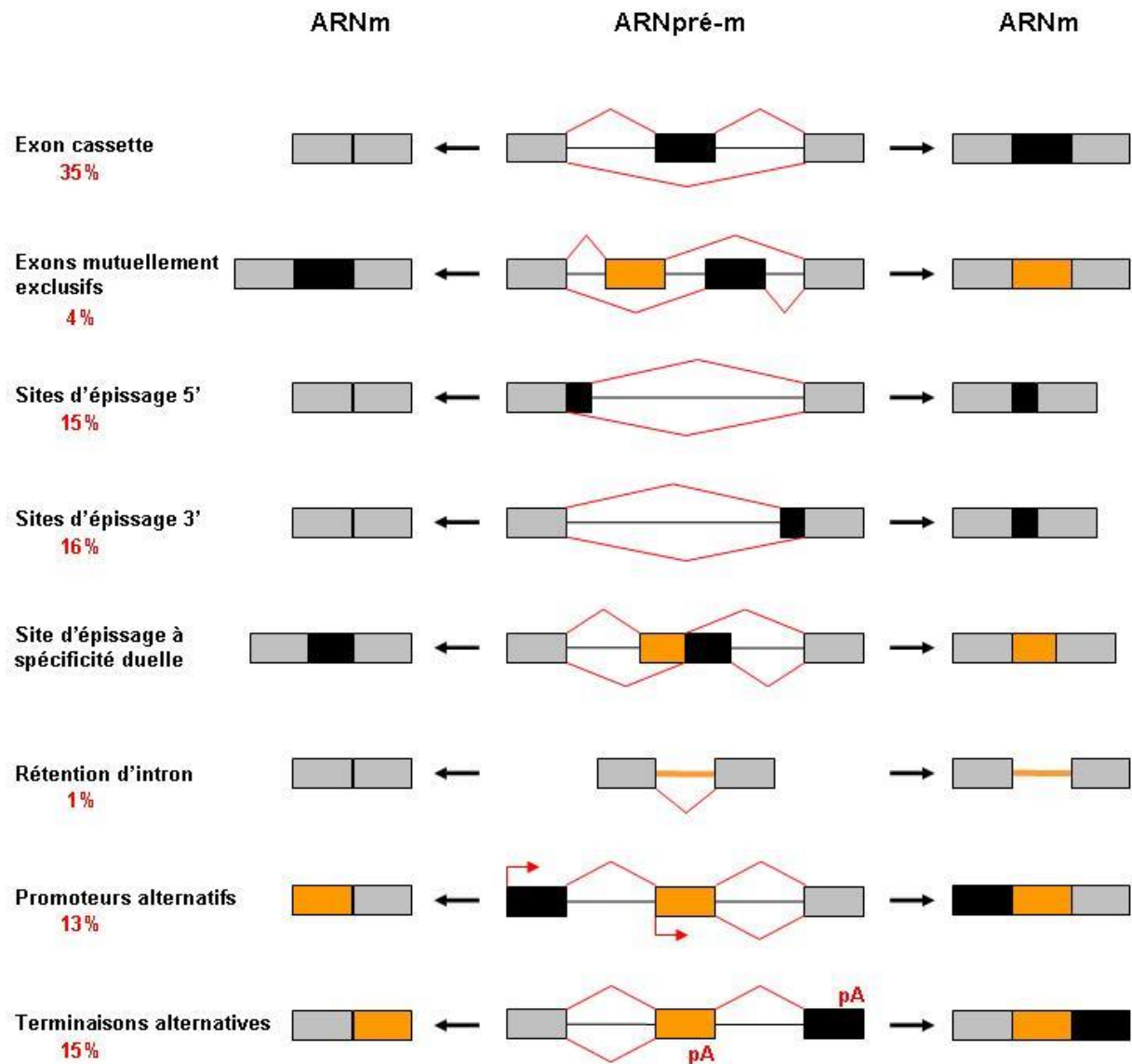


Figure 9. Les classes d'événements alternatifs menant à l'expression d'ARNm différents. Les pourcentage estimés de chaque types d'événements sont indiqués en rouge (Wang et al., 2008).

3.3. Les déterminants de l'épissage alternatif

Le choix des sites d'épissage 5' ou 3' détermine la séquence de l'ARNm produit. La sélection des sites d'épissage chez les eucaryotes est le résultat de l'influence de multiples facteurs, ce qui rend le déchiffrement et la compréhension de ce mécanisme extrêmement compliqués.

3.3.1. Les facteurs *cis*

a. Les sites d'épissage

Chez la levure, les sites d'épissage sont très conservés ce qui est aussi corrélé à une absence presque totale d'épissage alternatif. En revanche, chez les eucaryotes supérieurs, les sites d'épissage sont fortement dégénérés avec des séquences consensus faiblement établies ce qui n'est pas intrinsèquement suffisants pour contrôler l'assemblage du spliceosome. La reconnaissance et sélection de ces sites requière l'intervention de séquences régulatrices qui lient spécifiquement des facteurs en *trans*. Ainsi, la dégénération des sites d'épissage est proposée comme une composante importante du degré de régulation de l'épissage alternatif. (Wahl et al., 2009)

b. Les séquences régulatrices exoniques et introniques

Ce sont des séquences courtes, différentes des sites d'épissage 5' et 3', et qui modulent l'inclusion alternative des exons. Leurs séquences consensus sont aussi dégénérées. Elles peuvent avoir une activité stimulatrice de l'épissage et sont appelées ESE (*exonic splicing enhancer*) lorsqu'elles se trouvent dans les exons ou ISE (*intronic splicing enhancer*) lorsqu'elles se trouvent dans les introns. En opposition, celles induisant des effets opposés sont appelées ESS (*exonic splicing silencer*) ou ISS (*intronic splicing silencer*). Les facteurs protéiques se liant en *trans* sur ces éléments modulent le recrutement des snRNP du spliceosome.

Les ESE sont principalement liés par les protéines SR (c'est-à-dire contenant un domaine riche en Arginine-Sérine) alors que les ESS interagissent plutôt avec des protéines hnRNP. Ces deux types de protéines ont souvent été caractérisés pour leurs effets antagonistes sur le choix des sites d'épissage.

3.3.2. Les facteurs *trans*

L'ARN naissant est immédiatement pris en charge par de nombreuses protéines qui ont été historiquement nommées protéines hnRNP (pour *heterogenous ribonucleoprotein*). Elles portent souvent de multiples activités et par conséquent influencent profondément la maturation des ARN pré-m, notamment la régulation de l'épissage.

a. Les protéines SR

La famille des protéines SR comporte 9 membres canoniques ainsi que des facteurs à domaine RS apparentés (*SR-like*). Toutes (au moins les protéines canoniques) sont constituées d'un domaine RS, et d'un domaine RRM. (**Tableau 1**)

Name*	Domains	Binding sequence	Target genes
Canonical SR proteins			
SRp20 (SFRS3)	RRM and RS	GCUCCUCUUC	SRP20, CALCA and INSR
SC35 (SFRS2)	RRM and RS	UGCUGUU	ACHE and GRIA1–GRIA4
ASF/SF2 (SFRS1)	RRM, RRMH and RS	RGAAGAAC	HIPK3, CAMK2D, HIV RNAs and GRIA1–GRIA4
SRp40 (SFRS5)	RRM, RRMH and RS	AGGAGAAGGGA	HIPK3, PRKCB and FN1
SRp55 (SFRS6)	RRM, RRMH and RS	GGCAGCACCUG	TNNT2 and CD44
SRp75 (SFRS4)	RRM, RRMH and RS	GAAGGA	FN1, E1A and CD45
9G8 (SFRS7)	RRM, zinc finger and RS	(GAC)n	TAU, GNRH and SFRS7
SRp30c (SFRS9)	RRM, RRMH and RS	CUGGAUU	BCL2L1, TAU and HNRNPA1
SRp38 (FUSIP1)	RRM and RS	AAAGACAAA	GRIA2 and TRD
Other SR proteins			
SRp54	RRM and RS	ND	TAU
SRp46 (SFRS2B)	RRM and RS	ND	NA
RNPS1	RRM and Ser-rich	ND	TRA2B
SRrp35	RRM and RS	ND	NA
SRrp86 (SRrp508 and SFRS12)	RRM and RS	ND	NA
TRA2α	RRM and two Arg-rich	GAAARGARR	dsx
TRA2β	RRM and two RS	(GAA)n	SMN1, CD44 and TAU
RBM5	RRM and RS	ND	CD95
CAPER (RBM39)	RRM and RS	ND	VEGF

Tableau 1. La famille des protéines SR. (Chen and Manley, 2009)

Structure des protéines SR :

Le **domaine RS**, riche en sérine et arginine, lie l'ARNpré-m ainsi que d'autres facteurs à domaine RS. Ce domaine est la cible de nombreuses modifications post-traductionnelles. Il a cependant principalement été étudié dans le cadre de sa phosphorylation/déphosphorylation qui régule ses capacités d'interactions avec ses partenaires.

Le **domaine RRM** est un domaine de liaison à l'ARN. Il reconnaît des séquences spécifiques à chacune des protéines SR, principalement des ESE et ISE. Certaines protéines SR possèdent aussi un domaine pseudo-RRM, qui est proposé pour participer à des interactions protéines-protéines et stabilise l'interaction du RRM avec l'ARN.

Fonctions des protéines SR :

Les protéines SR modulent la sélection des sites d'épissage et leur activité est ciblée par la reconnaissance de séquences ESE. Par exemple, elles stabilisent la snRNP U1 au site d'épissage 5' et le facteur U2AF au site d'épissage 3'. Ces interactions passent par le domaine RS. (Chen and Manley, 2009) (**Figure 6D**). L'activité des protéines SR peut aussi être indépendante de séquences ESE (Hertel and Maniatis, 1999). Elles sont notamment impliquées dans l'incorporation au spliceosome de la tri-snRNP U4/U6-U5 (Roscigno and Garcia-Blanco, 1995).

Bien que l'activité stimulatrice des protéines SR soit la plus documentée, elles peuvent aussi avoir un rôle répressif, comme par exemple dans le cycle d'infection par l'adénovirus. La protéine SR, SF2/ASF, lie un ISS en amont du point de branchement de l'ARN de l'unité L1 de l'adénovirus s'opposant au recrutement de la snRNP U2 et à la sélection du site d'épissage 3' (Kanopka et al., 1996). Ainsi, selon le contexte, les protéines SR peuvent arborer deux rôles antagonistes.

b. Les protéines hnRNP

Les hnRNPs sont une vaste famille de facteurs dont certains membres ont été décrits pour arborer une activité régulatrice de l'épissage. (**Tableau 2**)

Name	Other names	Domains*	Binding sequences	Target genes
hnRNP A1	NA	RRM, RGG and G	UAGGGA/U	SMN2 and RAS
hnRNP A2	NA	RRM, RGG and G	(UUAGGG)n	HIV tat and IKBKAP
hnRNP B1				
hnRNP C1	AUF1	RRM	U rich	APP
hnRNP C2				
hnRNP F	NA	RRM, RGG and GY	GGGA and G rich	PLP, SRC and BCL2L2
hnRNP G	NA	RRM and SRGY	CC(A/C) and AAGU	SMN2 and TMP1
hnRNP H	DSEF1	RRM, RGG, GYR and GY	GGGA and G rich	PLP, HIV tat and BCL2L1
hnRNP H'				
hnRNP I	PTB	RRM	UCUU and CUCUCU	PTB, nPTB, SRC, CD95, TNNT2, CALCA and GRIN3B
hnRNP L	NA	RRM	C and A rich	NOS and CD45
hnRNP LL	SRRF	RRM	C and A rich	CD45
hnRNP M	NA	RRM and GY	ND	FGFR2
hnRNP Q	NA	RRM and RGG	ND	SMN2

Name	Other names	Domains*	Binding sequences	Target genes
hnRNP A1	NA	RRM, RGG and G	UAGGGA/U	SMN2 and RAS
hnRNP A2	NA	RRM, RGG and G	(UUAGGG)n	HIV tat and IKBKAP
hnRNP B1				
hnRNP C1	AUF1	RRM	U rich	APP
hnRNP C2				
hnRNP F	NA	RRM, RGG and GY	GGGA and G rich	PLP, SRC and BCL2L2
hnRNP G	NA	RRM and SRGY	CC(A/C) and AAGU	SMN2 and TMP1
hnRNP H	DSEF1	RRM, RGG, GYR and GY	GGGA and G rich	PLP, HIV tat and BCL2L1
hnRNP H'				
hnRNP I	PTB	RRM	UCUU and CUCUCU	PTB, nPTB, SRC, CD95, TNNT2, CALCA and GRIN3B
hnRNP L	NA	RRM	C and A rich	NOS and CD45
hnRNP LL	SRRF	RRM	C and A rich	CD45
hnRNP M	NA	RRM and GY	ND	FGFR2
hnRNP Q	NA	RRM and RGG	ND	SMN2

Tableau 2. La famille des hnRNP ayant un rôle dans la régulation de l'épissage. (Chen and Manley, 2009)

Bien que les hnRNP puissent agir de manière positive ou négative sur la sélection d'un exon, ces facteurs sont principalement documentés en tant qu'inhibiteur. Leur activité sur la reconnaissance des sites d'épissage peut se manifester par un blocage stérique des snRNP ou d'autres facteurs régulateurs. Comme les protéines SR, les hnRNP ciblent leurs activités par l'intermédiaire de séquences régulatrices. Par exemple, la protéine PTB (pour *polypyrimidine-tract binding protein*) est très affine pour l'étendue de polypyrimidine du site d'épissage 3' et interfère avec la liaison de U2AF (Singh et al., 1995). Ce mode d'action est décrit dans le cadre d'un modèle de « compétition ». D'autres modes d'action ont été proposés comme la multimérisation. Les hnRNP sont recrutées spécifiquement sur des séquences régulatrices et

initient leurs multimérisations locales sur l'ARN jusqu'à masquer des séquences activatrices ou des sites d'épissage. Un autre modèle propose un système de boucle dans laquelle l'exon alternatif serait « piégé » à l'intérieur de structures dépendantes des hnRNP (Chen and Manley, 2009).

c. Les autres régulateurs *trans* de l'épissage

Les facteurs d'épissage cités ci-dessus sont principalement exprimés de manière ubiquitaires. La famille des régulateurs de l'épissage est extrêmement élargie par ceux dont l'expression est spécifique d'un tissu ou d'un état cellulaire. Le cerveau est un très bon exemple avec notamment les facteurs NOVA, nPTB, nSR100/SRRM4 (Calarco et al., 2009; Irimia and Blencowe, 2012).

d. La dynamique et la flexibilité des facteurs d'épissage

La caractérisation de l'activité d'un facteur sur la régulation d'un exon alternatif ne suffit pas à prédire si cet exon sera inclus ou non. L'épissage est souvent contrôlé par des effets combinatoires et compétitifs de plusieurs facteurs d'épissage activateurs et répresseurs. Au final, c'est la concentration de ces facteurs et la somme de leur activité qui déterminent si un exon est inclus ou non (Mayeda et al., 1993; Zhu and Krainer, 2000). Cette capacité d'intégration dynamique du spliceosome joue un rôle essentiel dans les propriétés de modulation de l'épissage.

D'autre part, l'activité d'un facteur d'épissage peut varier en fonction de son positionnement relatif à l'exon. Ces facteurs se lient à l'ARN naissant avec une faible affinité en raison de la dégénérescence des séquences reconnues en *cis* (Shepard and Hertel, 2009). Plusieurs facteurs tels que NOVA1, NOVA2, FOX1 et FOX2 ont été montrés comme agissant en tant qu'activateurs ou répresseurs en fonction de leurs interactions en amont, en aval ou dans l'exon (Chen and Manley, 2009). Finalement les modifications post-traductionnelles sont souvent responsables d'une modulation de leurs interactions protéines-protéines ce qui participe à leur flexibilité notamment en réponse à divers stimuli.

3.3.3. Le cœur du spliceosome

Les composants du cœur du spliceosome sont aussi impliqués dans la régulation de l'épissage alternatif. Par exemple, l'équipe de B.J. Blencowe a mis en évidence que la réduction des niveaux des protéines SmB/B' par ARN interférence affecte l'épissage de nombreux exons alternatifs en comparaison des exons constitutifs (Saltzman et al., 2011). Le groupe de X.D. Fu a aussi montré que la régulation du niveau cellulaire de snRNP U2 active par la hnRNPU induit une perturbation de nombreux événements d'épissage alternatif (Xiao et al., 2012).

3.3.4. Les structures secondaires de l'ARNpré-m

La sélection des sites d'épissage peut aussi être influencée par des structures secondaires de l'ARN. D'une part, ces structures peuvent masquer les sites d'épissage (Grover et al., 1999; Vargas et al., 2011). D'autre part, elles peuvent interférer avec la fixation de facteurs d'épissage puisque ces derniers se lient à des séquences d'ARN simple brin (Hiller et al., 2007).

3.3.5. Les modifications post-traductionnelles

Des modifications post-traductionnelles amènent un degré supplémentaire à la dynamique du spliceosome. Ces modifications sont réversibles et mettent en jeu des enzymes qui apposent ou enlèvent ces modifications (phosphorylation, méthylation, acétylation, ubiquitination, glycosylation...). Elles influencent l'activité des composants du spliceosome, affectent l'assemblage du spliceosome, et agissent aussi sur la catalyse de l'excision des introns (Cheng et al., 2007; Friesen et al., 2001; Shepard and Hertel, 2009).

3.4. Quand la décision d'épissage alternatif est elle prise ?

Bien que, plusieurs études ont mis en évidence que la décision d'épissage peut être prise à différentes étapes de l'assemblage du spliceosome et même pendant les changements conformationnels entre les deux réactions de trans-estérifications (House and Lynch, 2006; Sharma et al., 2008), le choix des sites d'épissage se fait fréquemment à l'étape de leur

reconnaissance, c'est-à-dire au cours de l'assemblage précoce du spliceosome (complexe E notamment).

Aux vues de la flexibilité et du haut degré de régulation de l'épissage, il est important de déterminer la coordination spatio-temporelle de l'assemblage du spliceosome menant à la réaction d'épissage.

4. LE COUPLAGE DE L'ÉPISSAGE AVEC D'AUTRES PROCESSUS DE L'ARN

Dans cette partie, nous présenterons un nombre limité de couplages entre la maturation de l'ARN et l'épissage : la transcription, l'ajout de la coiffe, l'ajout de la queue poly A et l'édition et leurs rôles sur la régulation de l'épissage.

4.1. Le couplage de l'épissage avec la transcription

In vitro, la transcription et l'épissage sont deux processus qui peuvent être accomplies séparément. *In vivo*, ces deux processus s'effectuent de façon coordonnée dans le même compartiment : le noyau. Chez l'homme, la taille moyenne d'un gène est d'environ 10-15 kb mais elle peut atteindre plusieurs mégabases (Mb). Ainsi, la durée de la transcription des gènes peut atteindre plusieurs heures. Il semble donc difficilement concevable que la réaction d'épissage des premiers introns synthétisés débute seulement à la fin de la transcription. En effet, le gène de la dystrophine dont la longueur excède les 2 Mb requière environ 16 heures pour être transcrit et son épissage débute avant la fin de sa transcription (Tennyson et al., 1995).

4.1.1. L'épissage peut être co-transcriptionnel

La notion que l'épissage puisse être initié sur l'ARNpré-m en cours de synthèse a été introduite par Y.N. Osheim et A.L. Beyer à la fin des années 1980. Leurs travaux de microscopie électronique sur des embryons de drosophile montrent que les ARN naissants, encore attachés à la chromatine, sont chargés de particules ribonucléoprotéiques (Osheim et al., 1985). Ces particules sont localisées au niveau des jonctions exons-introns et sont donc proposées pour correspondre à la machinerie d'épissage. De plus, elles sont aussi observées

lors de la formation des boucles aux extrémités des introns et leur détection se réduit au fur et à mesure de la transcription (Beyer and Osheim, 1988) (**Figure 10**).

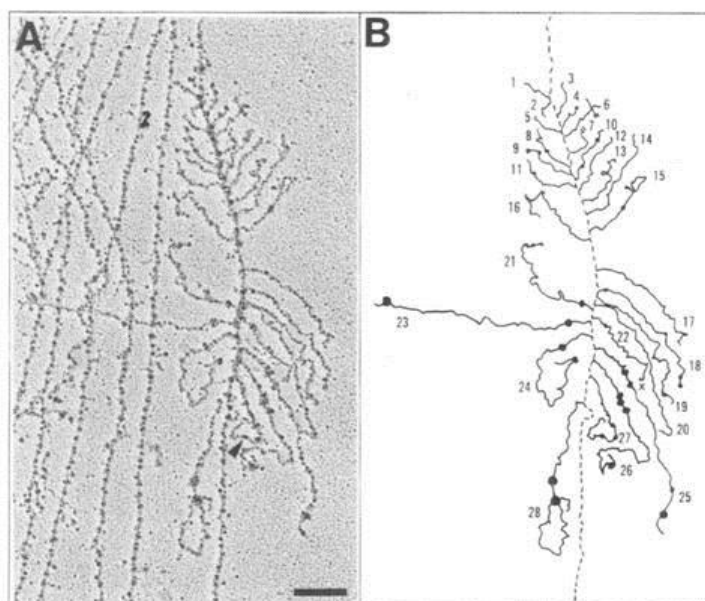


Figure 10. Formation de boucles sur les ARN naissants. **A.** Microscopie électronique d'une unité de transcription chez l'embryon de drosophile. La flèche indique une possible particule à un probable site d'excision d'intron. La barre représente 0,2 μ M. **B.** Représentation schématique. La ligne pointillée représente la chromatine ; les lignes pleines représentent la fibre RNP (ARN+hnRNP) ; les cercles pleins représentent les particules RNP et « x » représente les supposés transcrits cassés. Les transcrits sont numérotés de manière croissante du site d'initiation vers celui de terminaison.

L'isolement par fractionnement cellulaire des ARN associés à la chromatine *versus* ceux nucléoplasmiques et cytoplasmiques suggèrent que la plupart des introns sont épissés co-transcriptionnellement chez l'homme (Pandya-Jones and Black, 2009; Tilgner et al., 2012), la souris (Bhatt et al., 2012; Pandya-Jones et al., 2013) ou la drosophile (Khodor et al., 2011) et potentiellement chez la levure (Carrillo Oesterreich et al., 2010; Tardiff et al., 2006). L'utilisation du DRB (5,6-dichlorobenzimidazole 1- β -D-ribofuranoside), un inhibiteur réversible de la transcription, combinée à la quantification des ARNpré-m (Singh and Padgett, 2009) et le suivi de l'excision des introns par des techniques d'imagerie en temps réel (Martin et al., 2013) confortent l'idée que l'épissage est majoritairement catalysé co-transcriptionnellement.

Supportant cette idée, les protéines SR sont montrées pour co-localiser avec les sites de transcription active (Neugebauer and Roth, 1997; Zeng et al., 1997). L'expression transitoire d'un mini-gène contenant un intron provoque une relocalisation de la protéine SR ASF/SF2 depuis les *speckles* vers le site de transcription exogène (Misteli et al., 1997). Enfin, des expériences de ChIP (pour *Chromatin ImmunoPrecipitation* en anglais) suggèrent le

recrutement co-transcriptionnel des facteurs d'épissage (Görnemann et al., 2005; Lacadie and Rosbash, 2005; Listerman et al., 2006).

Comme nous l'avons évoqué précédemment, l'assemblage du spliceosome est dynamique, flexible et laisse place à de multiples mécanismes de régulation. L'excision des introns n'est pas sous contrôle d'un nombre limité de facteurs mais relate la somme d'une myriade d'activité. L'ensemble de ces processus sont autant d'étapes pendant lesquelles les interconnexions avec les autres machineries de maturation de l'ARN peuvent influencer l'épissage.

4.1.2. Le couplage transcription - épissage affecte l'épissage alternatif

Deux modèles, non exclusifs, ont été proposés pour expliquer les influences de la transcription sur la régulation de l'épissage alternatif.

Le modèle de recrutement, dans lequel il est proposé que l'ARN polymérase II interagit directement ou indirectement avec les composants du spliceosome pour favoriser ou défavoriser la réaction d'épissage.

Le modèle cinétique, met en jeu l'influence de l'élongation de la transcription sur l'inclusion d'exons alternatifs en perturbant la cinétique de recrutement du spliceosome et la synthèse de l'ARN naissant.

a. Le modèle de recrutement

Les arguments supportant ce modèle sont :

- Les promoteurs :

La modulation du contrôle de l'ARN polymérase II par différents promoteurs affecte l'épissage alternatif. L'équipe de A.R. Kornblihtt a montré que l'épissage de l'exon cassette EDI de la fibronectine est sujet à une régulation qui est dépendante du promoteur utilisé (Cramer et al., 1997). Les mêmes auteurs proposent aussi que les protéines SR, ASF/SF2 et 9G8, ont un recrutement différent sur l'exon EDI en fonction du promoteur (Cramer et al., 1999). Ces résultats supportent une connexion moléculaire entre le promoteur et l'exon alternatif EDI, et ces travaux suggèrent que les facteurs de transcription pourraient être un déterminant de cette régulation.

- Les facteurs de transcription :

L'activité de facteurs de transcription influence l'épissage alternatif (Allemand et al., 2008; Kadener et al., 2001, 2002; Rosonina et al., 2003). Les hormones stéroïdes perturbent l'épissage de mini-gènes rapporteurs lorsque ces derniers sont sous contrôle d'un promoteur sensible à ces hormones (Auboeuf et al., 2002). Il a été proposé que ce mécanisme fait intervenir des co-activateurs de la transcription : CAPER α et CAPER β (Dowhan et al., 2005). La sous-unité MED23 du complexe Médiateur a aussi été montré comme influençant l'épissage alternatif (Huang et al., 2012). MED23 interagit directement avec le facteur d'épissage hnRNP L et module la plupart des événements d'épissage alternatifs contrôlés par hnRNP L (Huang et al., 2012).

- L'ARN polymérase II :

Le développement d'un système *in vitro* de transcription et épissage a montré que l'ARN polymérase II participe à un couplage fonctionnel qui influence l'efficacité d'épissage (Das et al., 2006; Hicks et al., 2006). L'étude de ce couplage révèle que les protéines SR sont un élément essentiel et doivent être recrutées co-transcriptionnellement par l'ARN polymérase II (Das et al., 2007). La snRNP U1 est aussi proposée comme associée l'ARN polymérase II quand cette dernière est analysée par spectrométrie de masse (Das et al., 2007).

Le domaine carboxy-terminal (CTD) de la grande sous unité de l'ARN polymérase II a été suggéré comme un composant clef du couplage transcription - épissage. Chez l'homme, il est constitué de 52 répétitions de l'heptapeptide consensus YSPTSPS. L'évidence initiale de l'implication du CTD provient de l'observation qu'une version mutée de l'ARN polymérase II, sans CTD, n'est plus apte à épisser un ARN rapporteur (McCracken et al., 1997). Par ailleurs, le profil de phosphorylation du CTD affecte l'épissage (Hirose et al., 1999). La régulation du CTD affecte la régulation de l'épissage alternatif et cela semble se faire par l'intermédiaire de protéines SR (Muñoz et al., 2010; Rosonina and Blencowe, 2004). Des chromatographies d'affinité ont aussi mis en évidence que ce domaine interagit avec les protéines SR, U2AF, Prp40 (Hsin and Manley, 2012). Sur la base de ces observations, le domaine CTD est modélisé comme une plate-forme de recrutement de facteurs d'épissage. Néanmoins, le CTD, seul, n'est pas suffisant pour résumer les effet de l'ARN polymérase II sur l'épissage puisque couplé à la polymérase T7, il n'affecte pas l'épissage (Natalizio et al., 2009).

b. Le modèle cinétique

Modèle :

Initialement, le modèle cinétique a été établi sur l'évidence qu'une l'élongation lente de la transcription favorise l'inclusion d'un exon alternatif. Une vitesse élevée de l'élongation de l'ARN polymérase II génère la coexistence possible de plusieurs sites d'épissage. Cette coexistence établie une compétition entre des sites d'épissage de « force » différente et il en résulte de l'épissage alternatif. A l'inverse, une vitesse d'élongation lente tend à laisser un temps suffisant au spliceosome pour catalyser l'élimination de l'intron et abolit la possibilité de compétition entre sites d'épissage. Sur la base de ce modèle, il a été proposé qu'un exon alternatif arborant des sites d'épissage faible soit plus fréquemment inclus quand la vitesse d'élongation est faible. **(Figure 11A)**

Les évidences en faveur de ce modèle :

Ce modèle cinétique est largement inspiré d'expériences réalisées à l'aide de mutants de l'ARN polymérase II dont les vitesses d'élongation sont différentes. Chez la drosophile, la mutation C4 du CTD engendre une diminution de la vitesse d'élongation d'environ 2 fois, *in vitro* (Coulter and Greenleaf, 1985). Une analyse à grande échelle montre que l'épissage co-transcriptionnel est augmenté dans les cellules S2 de drosophile exprimant le mutant C4 (Khodor et al., 2011). Chez l'homme, le mutant R749H de la sous unité Rbp1, qui génère une mutation équivalente à la mutation C4, provoque aussi une différence de taux d'élongation d'environ 2 fois, *in vivo* (Boireau et al., 2007). Les travaux du groupe de A.R. Kornblihtt montrent que l'expression de la version humaine du mutant C4 favorise l'inclusion de l'exon alternatif EDI de la fibronectine (FN1), et celui d'E1a de l'adénovirus (de la Mata et al., 2003). Ces données ne permettent pas cependant d'exclure que la modulation observée sur l'épissage alternatif ne soient pas la conséquence d'effets indirects (changement de l'expression des composants du spliceosome, des facteurs d'épissage, recrutement différents de facteurs *via* l'ARN polymérase II...). En revanche, l'utilisation de ces mutants ne semble pas physiologiquement aberrante. La caractérisation de la vitesse d'élongation de 9 gènes suggèrent que celle-ci pourraient varier jusqu'à un maximum de 6 fois (Singh and Padgett, 2009).

L'utilisation de drogues inhibitrices conforte aussi la corrélation proposée entre la réduction de la vitesse d'élongation et l'augmentation de l'inclusion des exons alternatifs (Ip et al., 2011; Nogués et al., 2003). En accord avec cela, les travaux *in vitro* du groupe M. Ares

montrent une augmentation de l'inclusion de l'exon alternatif E2 de DYN2 avec des mutants « lents » de l'ARN polymérase II (Howe et al., 2003).

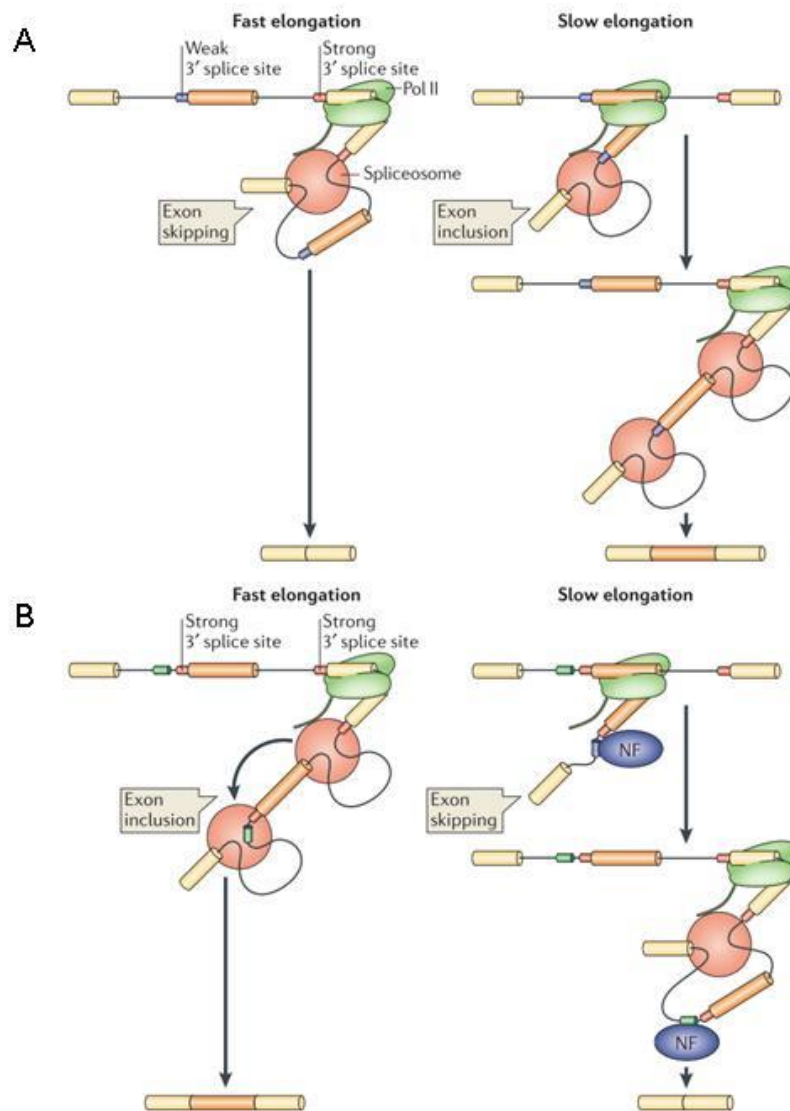


Figure 11. Modèle cinétique du couplage transcription - épissage. (Kornblihtt et al., 2013)

Exemples de facteurs protéiques modulant la vitesse d'élongation :

La protéine CTCF, une protéine qui se lie aux séquences d'ADN CCCTC, favorise l'inclusion de l'exon alternatif 5 du gène CD45. Il est proposé que l'interaction de CTCF dans l'exon 5 crée un « obstacle » induisant une pause de l'ARN polymérase II (Shukla et al., 2011).

Le complexe DBIRD, un co-facteur de l'ARN polymérase II, favorise l'exclusion d'exons contenant des séquences riches en A/T. Ces séquences sont particulièrement difficiles à transcrire et le complexe DBIRD est proposé pour faciliter l'élongation de la transcription sur ces séquences (Close et al., 2012).

Modèle revisité :

Le modèle évoqué jusqu'alors a récemment été revisité afin de proposer une corrélation entre la vitesse d'élongation et l'assemblage du spliceosome précoce. Ainsi, si un exon alternatif est régulé par un facteur répresseur, une élongation transcriptionnelle lente favorisera l'assemblage de complexe contenant ce répresseur (**Figure 11B**).

4.1.3. Qu'est ce que l'épissage co-transcriptionnel ?

La définition de l'épissage co-transcriptionnel est sujette à de multiples interprétations et ces dernières influencent fondamentalement la compréhension des mécanismes qui y sont rattachés.

L'épissage co-transcriptionnel peut être défini comme l'excision des introns catalysés au cours de la synthèse de l'ARNpré-m (Bhatt et al., 2012; Khodor et al., 2011; Pandya-Jones and Black, 2009; Pandya-Jones et al., 2013). Mais il est aussi envisageable qu'il réfère à l'assemblage co-transcriptionnel des complexes du spliceosome, lesquels dirigeront la réaction d'épissage une fois le transcrit relâché de la matrice ADN. Cette dernière possibilité déconnecte physiquement la réaction d'épissage du lieu de transcription.

Ces deux caractérisations de l'épissage co-transcriptionnel sont toutes deux supportées expérimentalement pour affecter l'épissage. Cependant, les conclusions mécanistiques qui peuvent être interprétées sont dépendantes de la définition choisie. Par conséquent, la définition de couplage co-transcriptionnel est encore un point de polémique.

Ainsi, pour faire transparaître l'aspect régulateur de la transcription sur l'épissage, nous proposons une définition : l'épissage est dit co-transcriptionnel si l'assemblage de la machinerie d'épissage est initié lorsque l'ARNpré-m est en cours de synthèse. Les travaux du groupe d'A.R. Kornblihtt montrent qu'une vitesse élongation plus lente favorise l'inclusion de l'exon EDI mais n'affecte pas l'ordre d'excision des introns. Ces données ont amené à revisiter le modèle du couplage co-transcriptionnel et proposer que la formation du complexe d'engagement du spliceosome en cours de transcription dirige la régulation de l'épissage alternatif (de la Mata et al., 2010). Les introns flanquant des exons alternatifs n'ont pas pour obligation d'être excisés pendant la synthèse de l'ARNpré-m (Ameur et al., 2011; Bhatt et al., 2012; Pandya-Jones and Black, 2009).

4.2. Le couplage de l'épissage avec d'autres processus de maturation des ARN

Les autres processus de maturation de l'ARNpré-m sont aussi décrits pour réguler l'épissage.

4.2.1. Le couplage de l'épissage et de la coiffe en 5' des ARNpré-m

Le modèle de définition de l'exon soulève des interrogations quant à l'épissage du premier intron de l'ARN naissant, en effet cet exon n'est défini que par un site d'épissage 5'. Il est proposé que l'excision du premier intron soit couplée à la maturation de la coiffe des ARNpré-m.

La première évidence du couplage entre l'ajout de coiffe et l'épissage provient de l'équipe de P.A. Sharp qui a montré avec des analogues du m⁷G de la coiffe, qu'un ARN composé de deux exons n'est épissé que si cet ARN est coiffé (Konarska et al., 1984). Par la suite, J.D. Lewis et ses collègues ont montré que la coiffe est impliquée dans le recrutement de la snRNP U1 au site d'épissage 5' du premier intron (Lewis et al., 1996)

4.2.2. Le couplage de l'épissage avec la maturation 3' terminale des ARNpré-m

Une problématique identique à celle ci-dessus est présente dans le contexte du dernier intron, en effet ce dernier n'est défini que par l'unique site d'épissage 3'. Il est proposé que la maturation terminale en 3' du transcrit participe à l'excision du dernier intron.

Des mutations du site de polyadénylation affectent considérablement l'efficacité d'épissage du dernier intron (Cooke et al., 1999; Niwa and Berget, 1991). De plus, les travaux de l'équipe de W. Keller montrent que le factor CPSF (*Cleavage and Polyadenylation Specifity Factor*) stimule l'excision du dernier intron en interagissant avec la snRNP U2 (Kyburz et al., 2006). L'inverse est aussi vrai, c'est à dire que les facteurs de l'épissage ont aussi un impact sur le traitement terminal du transcrit en 3' (Cooke et al., 1999; Kyburz et al., 2006).

Le couplage entre l'épissage et la finalisation du transcrit en 3' est suggéré comme un point de contrôle. Les facteurs de ces 2 machineries ont des interconnexions fonctionnelles qui les rendent nécessaires. En cas d'absence de l'une des deux machineries, les 2 processus sont inhibés.

4.2.3. Le couplage de l'épissage avec l'édition

L'édition des ARN est un processus qui effectue des changements post-transcriptionnel de la séquence de l'ARN. Il peut s'agir d'une insertion, d'une délétion ou de la conversion d'un nucléotide (Knoop, 2011).

Chez les mammifères, le processus majeur d'édition des ARN concerne la déamination d'une adénosine (A) en inosine (I). Cette réaction biochimique est catalysée par les enzymes ADAR (Knoop, 2011) qui reconnaissent des structures d'ARN double brins (Rieder and Reenan, 2012).

L'édition des ARN et l'épissage se déroulent dans le noyau après (ou en cours de) la transcription ; un couplage entre ces processus est désormais établi.

L'édition des ARN peut moduler l'épissage en créant de nouveaux sites d'épissage. Un exemple notable est celui de l'ARN Adar2 chez le rat, lequel est édité par la protéine ADAR2 pour convertir un dinucléotide AA en un dinucléotide AI (reconnu comme AG) créant par la même un nouveau site d'épissage 3' (Rueter et al., 1999).

Une édition en cours de transcription a été montrée pour réguler la formation de structure en double brins d'ARN. Ce type de modification peut masquer des jonctions exon-intron et moduler la reconnaissance de sites d'épissage (Rieder and Reenan, 2012). Les résultats du groupe de R.A. Reenan suggèrent que cette machinerie est couplée à l'activité d'hélicases qui peuvent déstructurer l'ARN et rétablir l'accessibilité des sites d'épissage (Reenan et al., 2000).

Par conséquent, l'édition des ARN peut influencer sur l'efficacité d'épissage en substituant ou masquant des séquences réactives ou régulatrices de l'épissage. Il est probable que la régulation de ce couplage se base sur le recrutement compétitif des facteurs impliqués dans l'édition et l'épissage.

Les activités de couplage dans la maturation des ARN se sont récemment complexifiées par l'incrémentation du nombre d'interconnexion entre les machineries. La transcription intervient dans la régulation de l'épissage. La chromatine et la transcription sont liées physiquement et fonctionnellement, par conséquent la chromatine est aussi proposée pour participer dans la modulation de l'épissage. Dans le chapitre suivant nous présenterons succinctement la chromatine, sa flexibilité structurale, sa composition, et son rôle fonctionnel avant d'explicitier dans le dernier chapitre les interconnexions entre la machinerie de l'épissage et la chromatine.

CHAPITRE 2 – LA CHROMATINE

1. LA STRUCTURE DE LA CHROMATINE

1.1. La compaction de la chromatine

Chez l'homme, le génome est composé de 3 milliards de paires de bases (pb). Mis bout à bout, toutes les molécules d'ADN de nos 23 paires de chromosomes constitueraient une fibre longue de 2 m qui est abritée dans le compartiment nucléaire mesurant environ 10 μ m. Cette observation amène un problème physique : comment le noyau réussit-il à intégrer un génome environ 200 000 fois plus grand que lui-même ? Ceci n'est possible qu'en imposant de multiples degrés de compaction à l'ADN. Le premier degré de repliement consiste à enrouler l'ADN autour d'un octamère de protéines nommées histones. Cette structure subit elle-même plusieurs degrés de compaction modifiant son degré de compaction.

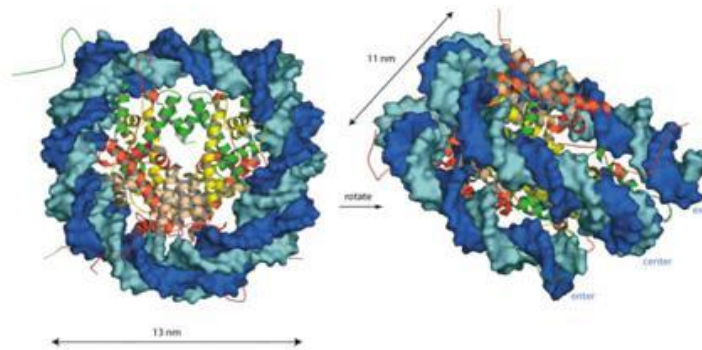
1.2. L'unité de base : le nucléosome

Chez les eucaryotes, l'unité de base de la chromatine est le nucléosome. Le nucléosome est composé d'un octamère d'histones, comprenant deux molécules de chacun des quatre histones H2A, H2B, H3 et H4, autour duquel s'enroule sur 1,7 tours un fragment d'ADN de 146 ± 2 pb (**Figure 12A**). Les octamères se forment de manière hiérarchique par l'assemblage deux à deux des histones: H2A avec H2B et H3 avec H4, ces derniers formant eux-mêmes un tétramère (H3-H4)₂, (**Figure 12B**). L'ADN est en contact avec le nucléosome avec une périodicité d'environ 10 pb. Par ailleurs, un cinquième type d'histone, l'histone H1 peut facultativement s'ajouter au nucléosome. Cette histone, appelée histone *linker*, effectue un lien physique entre les nucléosomes et l'ADN internucléosomique. L'histone H1 est impliquée dans la structure compacte et la stabilité de l'organisation des nucléosomes. (Khorasanizadeh, 2004)

Bien que la chromatine soit très condensée à l'échelle du noyau, localement elle arbore différents degrés de compaction. Dans chaque cellule, la chromatine abrite toute l'information génétique et doit donc être capable de libérer cette information (transcription, réplication, réparation de l'ADN...) dans un contexte spatio-temporelle adéquat. L'expression de

l'information génétique met en jeu de multiples régulations de la chromatine ciblant son état de relaxation local et le recrutement de nombreux effecteurs non histones. Dans les prochains paragraphes, nous allons présenter les différents processus de régulation de la chromatine.

A



B

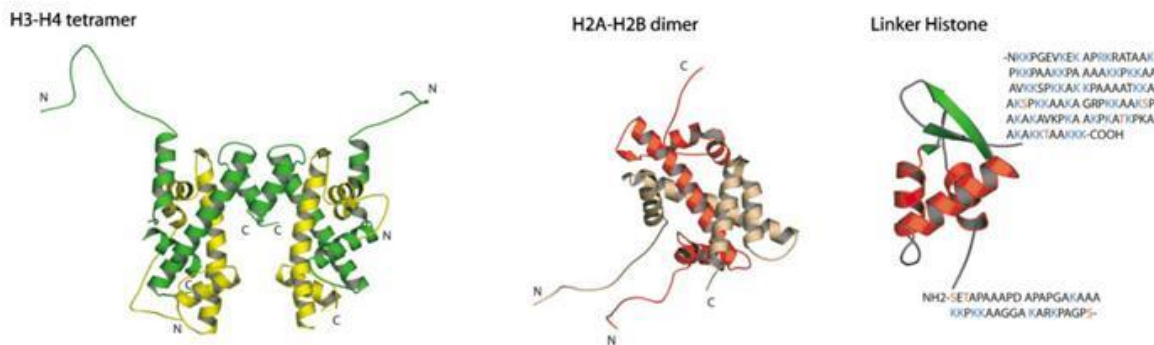


Figure 12. Le nucléosome. A. La structure atomique d'une particule nucléosomale. La double hélice d'ADN est représentée en bleu et les histones sont colorées comme en (B). B. La structure atomique des histones. (Khorasanizadeh, 2004)

2. LE REMODELAGE DE LA CHROMATINE

Les nucléosomes par eux-mêmes sont stables et peu mobiles pourtant la chromatine est une structure très dynamique. Ses propriétés dynamiques proviennent de l'action de complexes de remodelage qui restructurent, rendent mobiles ou éjectent les nucléosomes pour réguler l'accès à l'ADN. Ainsi, ces complexes sont impliqués dans tous les processus nucléaires impliquant l'ADN tels que la transcription, la réplication, la réparation...

Il existe plusieurs types de complexes de remodelage de la chromatine. Tous nécessitent l'hydrolyse de l'ATP pour leur fonctionnement. Tous les eucaryotes comprennent au moins 5 familles de remodeleurs de la chromatine : SWI/SNF, ISWI, NURD, INO80 et SWR1. Nous prenons l'exemple du complexe SWI/SNF (*mating-type SWItching – Sucrose Non Fermenting*) sur lequel nous reviendrons dans le cadre du couplage épissage - chromatine. (Saha et al., 2006)

SWI/SNF est un complexe multi-protéique dépendant de l'ATP. Il existe plusieurs complexes SWI/SNF qui diffèrent de part leurs composants. Notamment, SWI/SNF est composé des sous-unités catalytiques mutuellement exclusives BRM et BRG1. Les sous-unités catalytiques hydrolysent de l'ATP pour remanier les interactions entre l'ADN et les histones. Ces sous-unités catalytiques contiennent aussi un bromodomaine impliqué dans la liaison aux histones acétylées et par conséquent dans le ciblage de SWI/SNF à certains loci du génome. Le complexe SWI/SNF est principalement impliqué dans la réorganisation des nucléosomes pour faciliter le positionnement des facteurs de transcriptions, activateurs ou répresseurs. (Saha et al., 2006)

3. LES VARIANTS D'HISTONES

Les histones canoniques peuvent être substituées par des variants d'histones qui sont codés à partir de gènes différents. Ces variants ont le potentiel d'influencer considérablement la structure et les fonctions de la chromatine. Les variants d'histones sont retrouvés dans la plupart des organismes eucaryotes et sont exprimés dans tous les tissus à l'exception des variants d'histones H2B qui sont spécifiques des testicules. Par ailleurs, les variants d'histones arborent des degrés d'homologie en acides aminés s'étalant d'environ 40% à 99%. (Tableau 3)

Histone	Variant	Homologie	Taille (acides aminés)	Spécificité tissulaire	Localisation	Description Fonction
H2A	H2A (canonique)		129			
	H2A.X	82%	142		cassures double brins	réparation de l'ADN; H2A.X nécessite une phosphorylation dans la région C-terminale, la version phosphorylée est nommée γ H2A.X et permet le recrutement et l'accumulation de facteurs de la réparation de l'ADN
	H2A.Z	60%	127		hétérochromatine constitutive ; enhancers ; promoteurs	maintien de l'hétérochromatine mais aussi activation transcriptionnelle, les rôles en apparence contradictoires peuvent s'expliquer par différentes modifications post-traductionnelles (acétylation <i>versus</i> ubiquitination)
	macroH2A	65%	355		préférentiellement sur le chromosome X inactif	les séquences N et C terminales sont homologues de l'histone H2A canonique à environ 60% ; la partie centrale, appelée domaine Macro, est de type non histone et est composé d'acides aminés basiques, inhibe la transcription ; empêche le remodelage de la chromatine ; inactivation du chromosome X
	H2A.Bbd	48%	114		exclus du chromosome X inactif ; colocalise avec les histones acétylés	région C-terminale divergente ; sa composition induit une structure plus relâchée du nucléosome et pourrait être impliqué dans l'activation transcriptionnelle
H2B	H2B (canonique)		125			
	TH2B		127			
	hTSH2B	85%	127	sperme		
	H2BFWT	45%	175			
H3	H3.1 (canonique)		135			
	H3.2	99%	135			1 seul acide aminé diffère : la cystéine en position 96 est remplacée par une sérine ; l'expression et l'incorporation de ce variant est dépendante de la réplication
	H3.3	96%	135			5 acides aminés diffèrent dont 4 dans le domaine globulaire de l'histone ; l'incorporation est indépendante de la réplication ; enrichis dans les gènes transcriptionnellement actifs
	H3.1t	97%	135			
	CENP-A	45%	139		centromères	formation des kinétochores
	H3.X	79%	146	Cerveau		absence de résidus telle que la sérine 10 et 28
	H3.Y	78%	135			
H4	H4 (canonique)		102			
H1	H1 (canonique)		179			
	11 variants					

Tableau 3. Variants d'histones décrits chez l'homme. Les informations sont issues de : (Ausió, 2006; Talbert and Henikoff, 2010)

4. LES MODIFICATIONS EPIGENETIQUES DES HISTONES

Les histones sont des protéines de petite taille, riches en acides aminés basiques. Chaque histone comporte un domaine globulaire central de nature hydrophobe et des extrémités, appelées queues d'histones, moins conservées et moins structurées. Le domaine globulaire et principalement les queues d'histones sont sujets à des modifications post-traductionnelles associées à la régulation de nombreux processus nucléaires tels que la transcription, la réplication, la réparation de l'ADN. (Kouzarides, 2007a)

4.1. La diversité des modifications post-traductionnelles d'histones

4.1.1. La diversité à l'échelle du résidu d'acide aminé

Les modifications post-traductionnelles sont une source majeure de la régulation de la chromatine. D'une part, parce qu'au moins 60 résidus peuvent être affectés par ces marques. D'autre part, une multitude de groupements chimiques apposés sur les histones ont été caractérisés. Jusqu'au début des années 2010, une dizaine de marques d'histones avaient déjà été répertoriées : méthylation, acétylation, phosphorylation, sumoylation, citrullination, ubiquitination, ADP ribosylation, propionylation, butyrylation, formylation et isomérisation des prolines. Mais l'utilisation d'approches protéomiques basées sur la spectrométrie de masse a mis en exergue l'existence de près de 70 autres modifications d'histones (Kouzarides, 2007a; Tan et al., 2011). Et il est fort probable que l'augmentation de la sensibilité des approches expérimentales mènera à l'identification de nouvelles marques d'histones. Enfin, la diversité des marques épigénétiques est accrue par le nombre variable de groupements chimiques apposés sur un même résidu. Par exemple, les lysines et les arginines peuvent porter un, deux ou trois groupements méthyles.

Les fonctionnalités potentielles associées à ces marques épigénétiques détermine le code des histones et son décodage représente un défi majeur de la biologie moléculaire d'aujourd'hui.

4.1.2. La diversité à l'échelle du nucléosome

Un même nucléosome abrite plusieurs modifications d'histones. Ainsi la diversité du code des histones est largement amplifiée lorsque l'on prend en compte les combinaisons des différentes marques présentes sur un même nucléosome.

Les modifications d'histones sont interconnectées en *cis*, c'est-à-dire que plusieurs marques cohabitent sur la même histone et peuvent agir dans le même sens ou dans le sens opposé. Par exemple, au laboratoire, il a été mis en évidence que la reconnaissance de la méthylation de la lysine 9 de l'histone H3 par la protéine HP1 α (protéine de l'hétérochromatine 1) peut être perturbée par la citrullination de l'arginine 8 juste à côté (Sharma et al., 2012).

Aussi, les modifications sur des histones différentes d'un nucléosome peuvent être liées fonctionnellement, elles sont interconnectées en *trans*. Par exemple, au niveau de l'*enhancer* de FOSL1, la phosphorylation de la serine 10 de l'histone H3 (H3S10ph) induit le recrutement de l'adaptateur 14-3-3 qui lui-même recrute une histone acétyltransférase catalysant l'acétylation sur la lysine 16 de l'histone H4 (H4K16ac). La double modification H3K9acS10ph/H4K16ac crée une plate-forme pour le recrutement de la protéine à bromodomaine BRD4 favorisant ainsi le recrutement du facteur d'élongation P-TEFb au promoteur de FOSL1 (Zippo et al., 2009).

L'étude des combinaisons des modifications d'histones reste complexe puisque les techniques expérimentales actuelles ne permettent pas aisément de distinguer si les deux mêmes histones au sein d'un nucléosome possèdent les mêmes modifications.

4.2. Les modificateurs d'histones

La plupart des marques d'histones sont réversibles ce qui confère une grande plasticité à la chromatine. De nombreuses enzymes appelées « modificateurs d'histones » sont impliquées dans l'écriture des marques épigénétiques et d'autres permettent l'élimination de ces marques (**Figure 13**).

lysine 9 de l'histone H3 peuvent être invalidées indépendamment sans conséquence phénotypique grave, contrairement à la double invalidation (Peters et al., 2001).

Dans certains cas, la spécificité de ces enzymes peut être influencée par d'autres facteurs. D'une part, des complexes dans lesquels les enzymes sont trouvées peuvent spécifier une préférence pour les nucléosomes plutôt que pour les histones libres. Par exemple, le corépresseur CoREST stimule l'association entre la déméthylase d'histone LSD1 et le nucléosome favorisant ainsi la déméthylation des lysine 4 des histones H3 (Lee et al., 2005). D'autre part, des protéines qui s'associent avec l'enzyme peuvent affecter sa sélection à un résidu. Par exemple, en interagissant avec des récepteurs androgènes, LSD1 déméthyle les lysine 9 des histones H3, une marque répressive, au niveau des promoteurs des gènes de réponse aux androgènes (Metzger et al., 2005).

4.3. Les mécanismes associés aux fonctions des modifications d'histones

Les modifications d'histones exercent leurs effets *via* deux principaux mécanismes. Le premier implique une influence directe sur la structure de la chromatine. Le second implique la régulation du recrutement d'effecteurs à la chromatine.

4.3.1. Perturbation de la structure de la chromatine

Les modifications d'histones peuvent affecter la structure chromatinienne en modifiant les contacts entre des nucléosomes adjacents ou en modifiant les contacts entre l'ADN et les histones.

Sur la base principalement d'étude *in vitro* l'acétylation des histones a été décrite comme étant la modification des histones ayant le plus de potentiel pour modifier la structure de la chromatine. En effet, l'acétylation des histones joue un rôle important pour établir un espacement régulier entre les nucléosomes sur une matrice d'ADN (Loyola et al., 2001). De plus, l'acétylation des lysines 16 des histones H4 (H4K16) affecte négativement la formation de la fibre de 30 nm et la génération de structures d'ordre supérieur (Shogren-Knaak et al., 2006). Ces observations sont corroborées *in vivo*, par l'observation des niveaux d'acétylation des histones dans le noyau. L'acétylation semble distinguer l'euchromatine,

transcriptionnellement active et plutôt relaxée, de l'hétérochromatine, transcriptionnellement inactive et condensée (Johnson et al., 1998).

Le transfert du groupement acétyle permet de neutraliser la charge positive portée par les résidus lysines. La neutralisation de ces charges positives pourrait ainsi réduire la force d'interaction entre les histones acétylés et l'ADN, et aussi entre des nucléosomes différents. La phosphorylation des queues d'histone aurait aussi un tel effet en apportant des charges électrostatiques répulsives. La méthylation des résidus lysine quant à elle n'a pas d'effet neutralisant, aussi l'antagonisme fonctionnel qui existe entre l'acétylation et la méthylation sur certains résidus peut s'expliquer au moins en partie par des effets physiques sur la structure de la chromatine.

4.3.2. Régulation de la liaison de facteurs de la chromatine

Les modifications d'histones servent essentiellement de sites spécifiques de reconnaissance de domaines protéiques différents portés par les nombreux facteurs régulateurs de l'expression génique (**Figure 14**). La méthylation est reconnue par les domaines *chromo-like* de la famille Royale (chromo, tudor, MBT) et par des domaines PHD ; l'acétylation est reconnue par les bromodomaines et la phosphorylation par un domaine 14-3-3 (Kouzarides, 2007a).

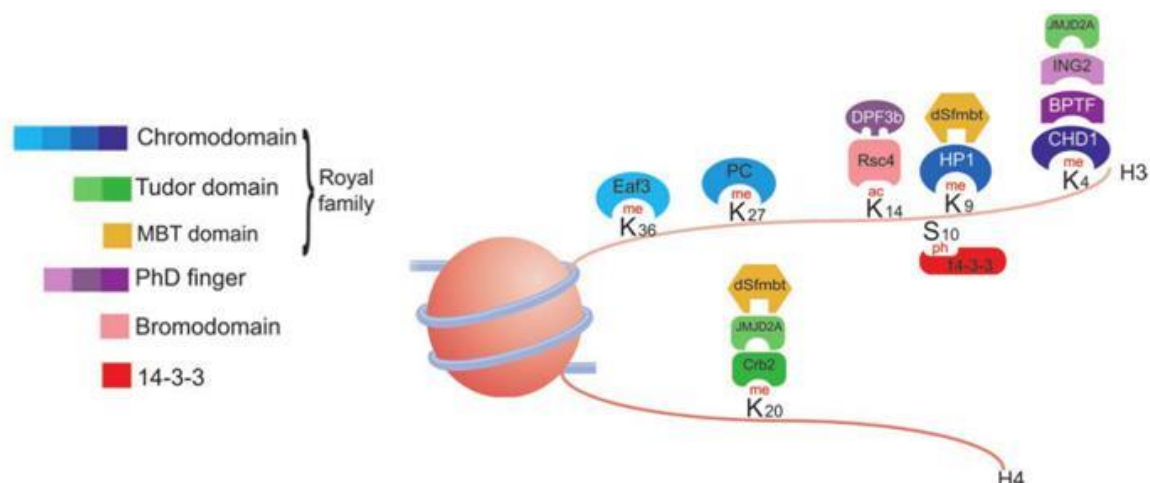


Figure 14. Domaines de liaison aux histones modifiées. (Bannister and Kouzarides, 2011)

Les « lecteurs » des modifications d'histones peuvent être associés à des modificateurs de la chromatine pour permettre leur action. Par exemple, les protéines HP1 qui reconnaissent la

méthylation des lysines 9 des histones H3 *via* leur chromodomaine, répriment l'expression du gène Survivin en interagissant avec la méthyltransférase d'ADN DNMT1 (Smallwood et al., 2007). Les lecteurs des marques d'histones peuvent aussi agir au sein de complexes enzymatiques. Par exemple, BPTF qui reconnaît H3K4me3 *via* son domaine PHD, appartient au complexe de remodelage de la chromatine NURF (type ISWI) et induit ainsi l'activation transcriptionnelle de HOXC8 (Wysocka et al., 2006). Enfin, les remodeleurs de la chromatine, tel que SWI/SNF possèdent des sous-unités catalytiques possédant des bromodomains de reconnaissance de l'acétylation qui permettent de cibler l'activité de mobilisation des nucléosomes lorsqu'ils sont acétylés.

De la même façon que les marques épigénétiques, leurs lecteurs peuvent aussi se combiner. Ainsi, deux modifications en *cis* sur la même queue d'histone peuvent servir à recruter deux lecteurs différents qui pourraient agir de concert. Ces deux modifications en *cis* pourraient aussi servir à recruter plus spécifiquement un lecteur qui posséderait les deux domaines de lecture. La combinaison des lecteurs pourrait aussi servir à reconnaître des modifications en *trans*, sur deux histones différentes d'un même nucléosome ou sur des nucléosomes différents (Musselman et al., 2012).

4.4. Les modifications d'histones et la transcription

Les approches de séquençage à haut débit ont permis de caractériser en détail les modifications d'histones associées aux différentes régions du génome. En particulier, des zones transcriptionnellement actives ou réprimées sont associées à des marques d'histones différentes. Les régions inactives sont principalement associées aux méthylations des lysines 9 et 27 des histones H3. Les gènes transcriptionnellement actifs sont associés à d'autres marques qui, de plus, diffèrent selon la région du gène. Les promoteurs des gènes actifs sont notamment reliés à la triméthylation de la lysine 4 de l'histone H3 alors que les corps des gènes sont corrélés à la présence des lysines 36 et 79 méthylées de l'histone H3. (Kouzarides, 2007a; Smolle and Workman, 2013) (**Figure 15**)

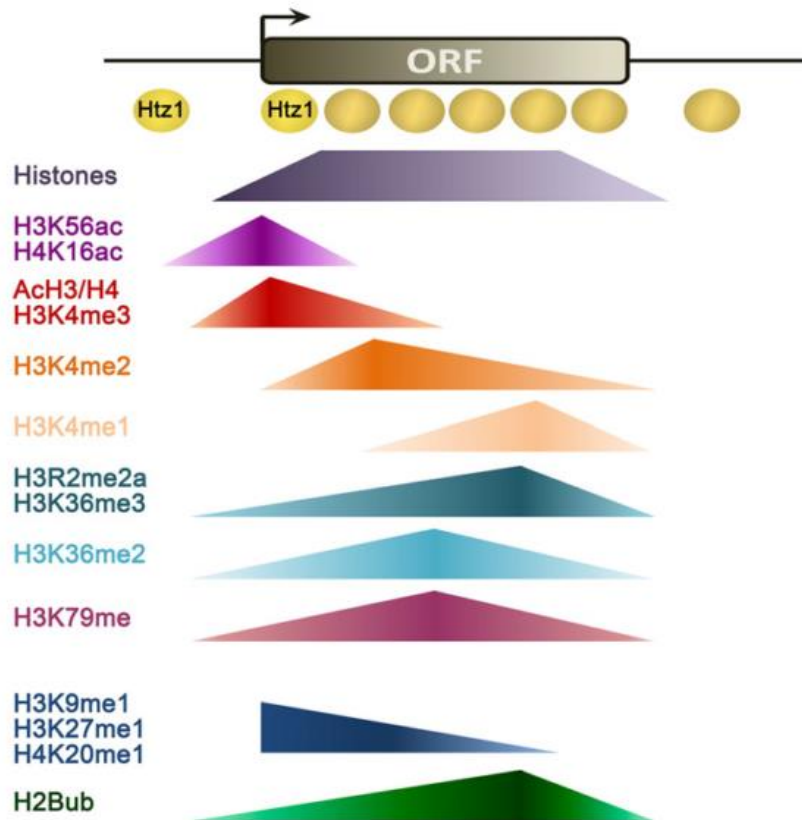


Figure 15. Profils des modifications d'histones associées à un méta-gène transcriptionnellement actif. (Smolle and Workman, 2013)

5. LES MODIFICATIONS EPIGENETIQUES DE L'ADN

5.1. La méthylation de l'ADN

La méthylation de l'ADN est un mécanisme clef dans la régulation épigénétique chez les eucaryotes. Les patrons de méthylation de l'ADN sont établis dans au moins deux périodes du développement : dans les cellules germinales et dans les embryons pré-implantatoires (Reik, 2007). Des mécanismes de maintien sont impliqués dans la conservation des profils corrects de méthylation de l'ADN dans les cellules somatiques. Par ailleurs, la méthylation de l'ADN joue un rôle crucial dans la régulation de la pluripotence, l'inactivation du chromosome X et dans l'empreinte génomique (Bird, 2002). La méthylation de l'ADN assure aussi l'intégrité et la stabilité du génome en inhibant les éléments transposables (Goll and Bestor, 2005).

5.1.1. Localisation de la méthylation de l'ADN

La méthylation de l'ADN est trouvée dans de nombreux organismes chez les procaryotes et les eucaryotes. Chez les procaryotes, la méthylation de l'ADN peut se faire sur les cytosines et les adénines. La méthylation d'ADN est majoritairement conservée chez la majorité des eucaryotes, notamment chez les plantes, les champignons et les animaux. En revanche, certains organismes ne possèdent pas ou très peu de méthylation de l'ADN comme par exemple, chez les levures et chez plusieurs invertébrés tels que la drosophile et le nématode *Caenorhabditis elegans*. Chez les eucaryotes la méthylation est presque exclusivement présente sur les cytosines.

a. A l'échelle nucléotidique

Chez les mammifères la méthylation se fait principalement sur les cytosines dans un contexte de dinucléotides CpG. Néanmoins, l'étude du méthylome avec une résolution d'un seul nucléotide a permis de mettre en évidence que dans les cellules souches embryonnaires humaines H1 près de 25% des cytosines méthylées se trouvent dans un contexte non CpG (Lister et al., 2009).

b. A l'échelle génomique

Chez les mammifères la proportion en dinucléotides CpG est faible : la fréquence observée est d'environ 0,8% alors que la fréquence attendue est de 4% compte tenu des proportions de C et G. La perte des dinucléotides CpG dans le génome serait due à la conversion des 5mC en thymidine à la suite d'une déamination (Bird, 1980).

Parmi les 28 millions de CpG, 60 à 80% sont méthylés. Moins de 10% des CpG se trouvent dans des régions denses en CG appelées îlots CpG. Ces îlots sont répandus dans les promoteurs des gènes de ménage et des gènes régulateurs du développement (Deaton and Bird, 2011). 70% des promoteurs sont riches en CpG (Saxonov et al., 2006) qui pour la plupart ne sont pas méthylés.

5.1.2. La méthylation de l'ADN est catalysée par les enzymes DNMT

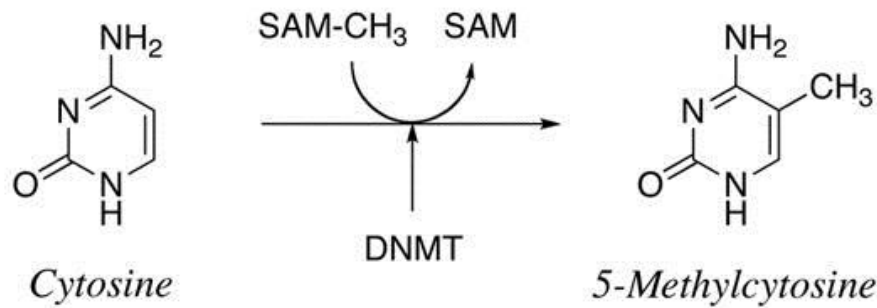


Figure 16. Conversion d'une cytosine en 5-méthylcytosine par une ADN méthyltransférase.

La méthylation de l'ADN consiste en l'ajout d'un groupement méthyle (CH₃) sur le carbone en position 5 de la cytosine (5mC). La réaction est catalysée par des ADN méthyltransférases (DNMT) qui utilisent le groupement méthyle (CH₃) de la S-adénosylméthionine (SAM). (**Figure 16**) Chez les mammifères, plusieurs protéines DNMT ont été caractérisées (**Figure 17**). Ces enzymes sont essentielles pour l'établissement et le maintien des profils de méthylation des cytosines.

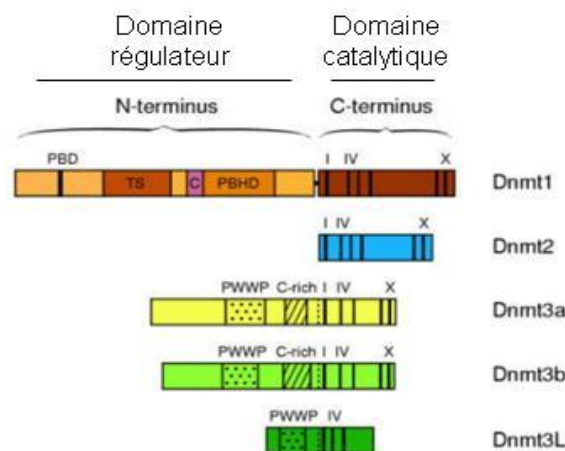


Figure 17. La famille des ADN méthyltransférases (DNMT).

a. DNMT1

DNMT1 a été la première ADN méthyltransférase décrite chez la souris (Bestor et al., 1988).

C'est principalement une méthyltransférase de maintien qui préserve les profils de méthylation au fil des divisions cellulaires. Cette enzyme agit de façon non spécifique d'un site donné en agissant au niveau des foyers de réplication pendant la phase S et en méthylant

préférentiellement les CpG hémi-méthylés. Ces CpG qui ne sont méthylés que sur un seul des deux brins d'ADN sont reconnus par le facteur UHRF1 (cf. paragraphe 5.1.4.c. de ce chapitre) dont l'interaction avec DNMT1 permet la méthylation des CpG nouvellement synthétisés (Avvakumov et al., 2008; Sharif et al., 2007).

DNMT1 est aussi être recrutée sur des loci spécifiques par des interactions avec d'autres co-facteurs, ce recrutement ayant pour effet de réprimer la transcription (cf. paragraphe 5.1.3. de ce chapitre) (Robertson et al., 2000). Il y a aussi des évidences que DNMT1 possède une activité *de novo* dans des cellules cancéreuses humaines (Jair et al., 2006).

L'invalidation du gène *Dnmt1* chez la souris provoque la létalité embryonnaire au jour 9,5 et est associée à une importante hypométhylation (perte des 2/3 de la méthylation) (Li et al., 1992). En revanche, en culture, les cellules souches embryonnaires (ES) issues des embryons *Dnmt1*^{-/-} sont viables, ne présentent pas de défauts de prolifération, de morphologie (Lei et al., 1996; Li et al., 1992) et arborent des niveaux affaiblis mais stables de méthylation d'ADN suggérant l'existence d'autres enzymes DNMT.

b. DNMT2

Des recherches par homologie de séquences avec le domaine catalytique de DNMT1 ont mené à la découverte de DNMT2. Cette protéine a longtemps été considérée comme une énigme parce qu'en dépit de sa forte homologie avec les autres enzymes DNMT elle ne semble pas posséder d'activité ADN méthyltransférase *in vitro* et *in vivo* (Durdevic and Schaefer, 2013).

En revanche, des approches génétiques et biochimiques ont permis de mettre en évidence que cette enzyme catalyse la méthylation de la cytosine 28 de l'ARNt de l'acide aspartique (Goll et al., 2006). DNMT2 n'est pas une DNMT mais plutôt une ARN méthyltransférase (Defossez Bioessay 2013).

c. La famille des DNMT3

La famille des DNMT3 est composée de trois protéines qui présentent de fortes homologies entre elles. Deux d'entre elles, DNMT3A et DNMT3B sont catalytiquement actives. En revanche, DNMT3L (*DNMT3-Like*) ne possède pas d'activité enzymatique mais est impliquée dans la régulation de la méthylation de l'ADN (Bestor, 2000).

DNMT3A et DNMT3B :

L'invalidation du gène *Dnmt3a* chez la souris provoque la létalité 4 semaines après la naissance tandis que l'invalidation de *Dnmt3b* conduit à la mort embryonnaire au jour 9,5 (Okano et al., 1999). Ceci démontre que les DNMT ne sont pas redondantes et peuvent avoir des rôles spécifiques. Les enzymes DNMT3A et DNMT3B sont responsables de la méthylation *de novo* pendant le développement embryonnaire et la gamétogénèse (Okano et al., 1999).

L'activité des enzymes DNMT3 n'est pas restreinte à la méthylation *de novo*. En effet, dans les cellules humaines de carcinome colorectal HCT116, l'invalidation de DNMT1 conduit à une réduction de seulement 20% de la méthylation globale du génome (Rhee et al., 2000). En revanche, la double inactivation de DNMT1 et DNMT3B conduit à une perte de plus de 95% de la méthylation (Rhee et al., 2002). Par ailleurs, DNMT3B semble aussi agir sur le maintien de la méthylation de l'ADN (Velasco et al., 2010).

DNMT3L :

Bien que DNMT3L ne présente aucune activité méthyltransférase, cette protéine est indispensable pour la méthylation *de novo* de la plupart des loci à empreinte parentale dans les cellules germinales (Jia et al., 2007). DNMT3L stabilise la conformation du site actif de DNMT3A et stimule la liaison de la S-adénosylméthionine (Jia et al., 2007).

5.1.3. La méthylation de l'ADN est impliquée dans la répression transcriptionnelle

La méthylation de l'ADN au promoteur est impliquée dans la répression transcriptionnelle des gènes associés. Plusieurs modèles peuvent expliquer le lien entre ces deux phénomènes.

La méthylation de l'ADN pourrait agir directement sur la répression des gènes en empêchant le recrutement d'activateurs transcriptionnels à leur site de liaison (Watt and Molloy, 1988).

Par ailleurs, la répression permise par la méthylation de l'ADN est plus efficace dans un contexte chromatinien (Buschhausen et al., 1987; Kass et al., 1997) suggérant une coopération entre la méthylation de l'ADN et les histones. Notamment, les enzymes DNMT interagissent avec des protéines impliquées dans des modifications d'histones associées à la répression transcriptionnelle. Par exemple, les DNMT interagissent avec SUV39H1 et G9A des enzymes

qui méthylent la lysine 9 de l'histone H3 et avec les protéines de l'hétérochromatine HP1 (Estève et al., 2006; Fuks et al., 2003; Geiman et al., 2004; Smallwood et al., 2007). Les enzymes DNMT sont aussi impliquées dans la répression transcriptionnelle *via* leur interaction avec des histones déacétylases (Geiman et al., 2004; Robertson et al., 2000).

Enfin, des protéines capables de se lier à l'ADN méthylé, que nous allons présenter dans le paragraphe suivant, peuvent aussi faire le lien entre la méthylation d'ADN et la répression transcriptionnelle.

5.1.4. Les lecteurs de la méthylation de l'ADN jouent un rôle important dans l'action répressive de la méthylation de l'ADN

a. La famille des protéines à domaine MBD

Basé sur leur capacité à se lier à l'ADN méthylé *in vitro*, les membres de la famille des protéines à domaine de liaison au CpG méthylés (MBD) sont les principaux lecteurs de la méthylation de l'ADN (Hendrich and Bird, 1998; Meehan et al., 1989). La famille des protéines à domaine MBD (*methyl-CpG-binding domain*) comporte 5 membres : MeCP2, MBD1, MBD2, MBD3 et MBD4. Toutes, excepté MBD3 sont capables de se lier à l'ADN méthylé. MBD3 ne peut pas se fixer aux CpG méthylés *in vitro* en raison de mutations dans son domaine MBD (Saito and Ishikawa, 2002; Zhang et al., 1999). **(Figure 18)**

A l'exception de MBD3 qui est essentielle au développement embryonnaire (Hendrich et al., 2001), les invalidations de Mbd1, Mbd2 et Mbd4 chez la souris (Hendrich et al., 2001; Martín Caballero et al., 2009) ont des conséquences légères sur le développement suggérant des fonctions redondantes des protéines MBD. Des mutations dans MeCP2 provoquent le syndrome de Rett qui se manifeste par des désordres neurologiques chez les femelles et la mort chez les mâles (Amir et al., 1999).

Des études d'interaction protéiques et des approches *in vitro* ont mené à l'élaboration d'un modèle dans lequel MBD1, MBD2, MBD4 et MeCP2 recrutent des remodeleurs de la chromatine, des histones déacétylases et méthylases au locus de l'ADN méthylé pour induire une répression transcriptionnelle (Kondo et al., 2005; Nan et al., 1998; Ng et al., 1999). MBD3, même sans capacité à lier à l'ADN méthylé participe à la répression transcriptionnelle car c'est un composant du co-represseur NuRD, un complexe de remodelage et de déacétylases d'histones (Zhang et al., 1999).

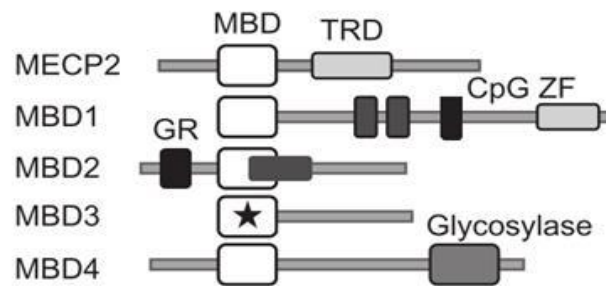


Figure 18. La famille des protéines à domaine MBD. MBD, *methyl-CpG-binding domain* ; TRD, *transcriptional repression domain* ; CpG ZF, *unmethylated-CpG-binding zinc finger* ; glycosylase, *DNA glycosylase activity*. MBD3 a un domaine MBD qui ne lie pas l'ADN méthylé (indiqué par une étoile). (Defossez and Stancheva, 2011)

b. Les protéines Kaiso et Kaiso-like

Les protéines Kaiso appartiennent à la famille des protéines à doigts de zinc. Kaiso se lie préférentiellement, *via* ses domaines à doigts de zinc, à des CpG méthylés dans certains contextes (Prokhortchouk et al., 2001; Sasai et al., 2010). Cependant, ces protéines ont aussi été retrouvées liées à des séquences d'ADN non méthylées (Daniel et al., 2002; Sasai et al., 2010). Par l'intermédiaire de son domaine BTB (*Bromodomain transcription factors and PHD domain*), Kaiso peut inhiber la transcription en recrutant le co-represseur NCoR (Yoon et al., 2003). (**Figure 19**)

Par ailleurs, des recherches d'homologie ont permis d'identifier 2 protéines, ZBTB4 et ZBTB38, proches des protéines Kaiso qui partagent plusieurs de leurs propriétés (Filion et al., 2006). Elles peuvent se lier à l'ADN méthylé ou non (Sasai et al., 2010) et inhiber la transcription en recrutant le répresseur transcriptionnel CtBP (*C-terminal binding protein*) (Sasai et al., 2005). (**Figure 19**)

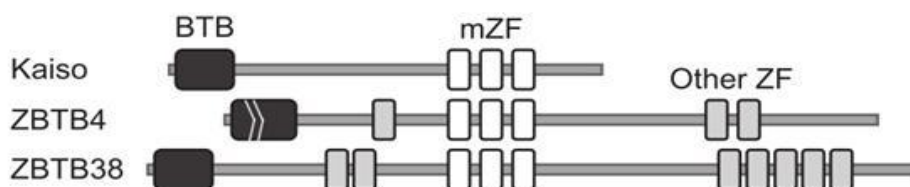


Figure 19. Les protéines Kaiso et Kaiso-like. BTB, *BTB/POZ domain* ; mZF, *methyl-CpG-binding zinc fingers*. (Defossez and Stancheva, 2011)

c. La famille des protéines à domaine SRA

Cette famille est composée de 2 membres, UHRF1 et UHRF2, qui présente une très forte homologie (**Figure 20**). Bien qu'UHRF1 ait été initialement décrite comme facteur de liaison à des séquences d'ADN non méthylées (Hopfner et al., 2000), leur domaine SRA leur confère la capacité à se lier à l'ADN méthylé.

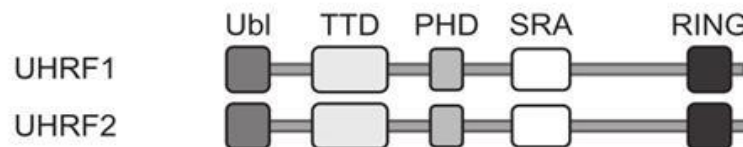


Figure 20. La famille des protéines à domaine SRA. Les 2 membres de la famille sont UHRF1 et UHRF2. Ubl, *ubiquitin-like domain* ; TTD, *tandem tudor domain* ; PHD, *PHD finger domain* ; SRA, *SRA domain* ; RING, *RING finger domain*. (Defossez and Stancheva, 2011)

UHRF1 :

UHRF1 (*ubiquitin-like PHD and RING finger domain-containing protein 1*) est parmi les protéines de liaison à l'ADN méthylé la seule qui est essentielle chez la souris. Son invalidation conduit à un arrêt du développement embryonnaire juste après la gastrulation et la méthylation globale des CpG est considérablement réduite (Sharif et al., 2007).

UHRF1 est une E3 ubiquitine ligase qui arbore plusieurs domaines fonctionnels dont un domaine Tudor qui lui permet de se lier à des modifications post-traductionnelles d'histones et un domaine SRA (*SET and RING-associated*) pour reconnaître la méthylation de l'ADN. UHRF1 reconnaît les cytosines méthylées dans un contexte de dinucléotides CpG hémiméthylés (Avvakumov et al., 2008). Des études génétiques chez la souris ont mis en évidence que UHRF1 est indispensable pour le maintien de la méthylation dépendante de DNMT1 (Sharif et al., 2007). Cette fonction nécessite la capacité d'UHRF1 à se lier à H3K9me2/3 (Rothbart et al., 2012).

Par ailleurs, UHRF1 peut agir comme un répresseur transcriptionnel en recrutant l'histone méthyltransférase G9A (Kim et al., 2009).

UHRF2 :

UHRF2, comme UHRF1 se lie à H3K9me grâce à son domaine Tudor et se lie à l'ADN méthylé grâce à son domaine SRA (Zhang et al., 2011). En revanche, UHRF2 ne semble pas impliquée dans le maintien de la méthylation de l'ADN dépendante de DNMT1 au cours des réplifications (Zhang et al., 2011).

5.2. La déméthylation de l'ADN

Bien que la méthylation de l'ADN dans les cellules somatiques soit stable, une perte globale de la méthylation de l'ADN est observée à différentes étapes du développement comme dans l'embryon préimplantatoire et le développement des cellules germinales primordiales. La déméthylation peut s'effectuer de deux façons :

- la **déméthylation passive** qui consiste à la perte de la méthylation au cours des réplifications successives en absence de machinerie de maintien de la méthylation fonctionnelle,
- la **déméthylation active** qui consiste à enlever ou modifier le groupement méthyle des cytosines par un processus enzymatique.
-

La déméthylation active a longtemps été controversée. Cependant de récentes évidences indiquent que ce processus de déméthylation active existe et qu'il fait intervenir des enzymes nommées TET.

La famille des enzymes TET est composée de 3 membres : TET1, TET2 et TET3. Ce sont des di-oxygénases qui catalysent l'oxydation de la méthylation des cytosines (5mC) en hydroxyméthylation (5hmC) (Ito et al., 2010; Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). Les TET sont capables d'oxydations itératives, c'est-à-dire qu'après avoir générer la 5hmC, les TET peuvent l'oxyder en 5-formylcytosine (5fC) puis en 5-carboxylcytosine (5caC) (He et al., 2011; Ito et al., 2011). La régénération d'une cytosine non modifiée (C) se fait soit par dilution passive dépendamment de la répllication, soit par l'intervention du système de réparation par excision de base (BER) (Kohli and Zhang, 2013; Wu and Zhang, 2014). **(Figure 21)**

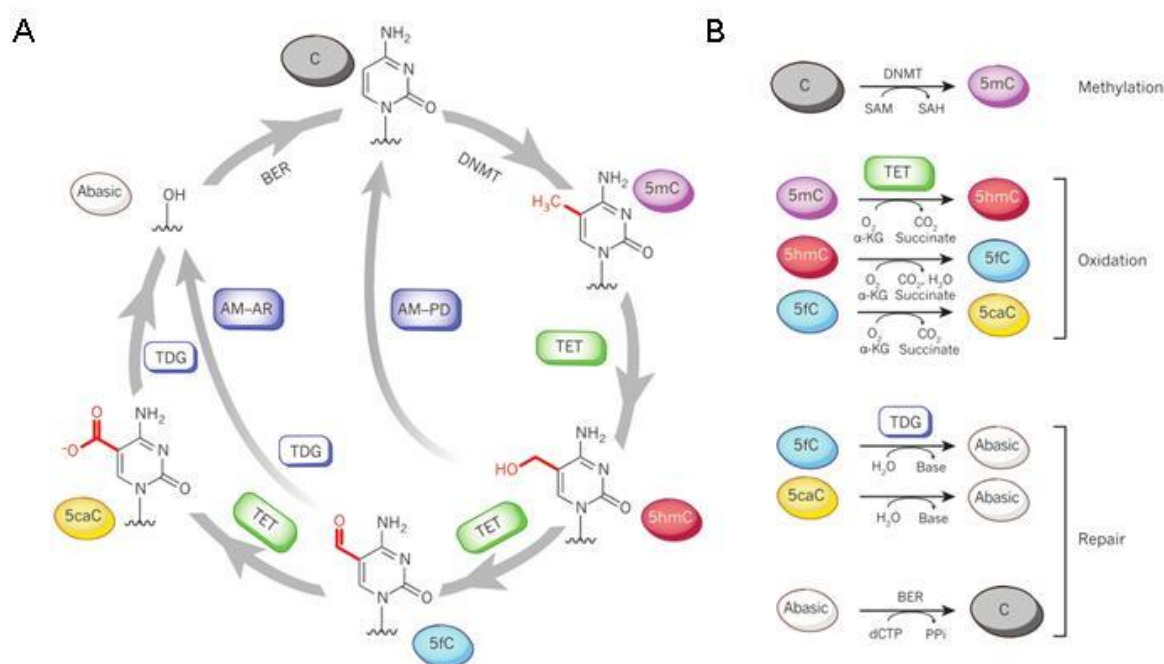


Figure 21. Les différentes voies de déméthylation actives de l'ADN. **A.** Les cytosines méthylées (5mC) par les enzymes DNMT peuvent être oxydées itérativement en 5hmC, 5fC et 5caC. Dans la voie de modification active (AM), suivie d'une dilution passive (PD), 5hmC est diluée avec la réplication de façon à régénérer une cytosine non modifiée C. Pour plus de clarté, les PD des composés hautement oxydés 5fC et 5caC ne sont pas représentées. Dans la voie de AM suivie par une restauration active (AR), 5fC et 5caC sont excisées par TDG générant un site abasique puis dans le cadre de la réparation par excision de base (BER) une cytosine non modifiée C est régénérée. **B.** Les réactions individuelles de la voie de déméthylation active sont indiquées à droite. (Kohli and Zhang, 2013)

6. LES ARN NON CODANTS

Il est maintenant connu que les petits ARN peuvent induire la répression des gènes de manière transcriptionnelle par un processus nommé TGS (*Transcriptional Gene Silencing*) qui implique une régulation épigénétique de la structure de la chromatine. Le TGS a d'abord été observé chez la plante, la drosophile et la levure et plus récemment observé chez l'homme (Han et al., 2007; Kim et al., 2006; Morris et al., 2004; Weinberg et al., 2006).

Chez la levure *S. pombe*, les ARN transcrits des régions répétées péricentromériques sont maturés par Dicer et chargés sur la protéine Argonaute 1 (AGO1). Ceci conduit au recrutement d'enzymes qui méthylent la lysine 9 de l'histones H3 (H3K9), et par conséquent favorise le recrutement de Swi6/HP1 pour aboutir à la formation de hétérochromatine (Grewal, 2010).

Plusieurs études suggèrent que lors du TGS induit par des siRNA, des protéines AGO seraient recrutées spécifiquement aux promoteurs des gènes cibles (Kim et al., 2006; Noma et al., 2004). Le TGS est associé à certaines marques épigénétiques telles que la méthylation des lysines 9 et 27 des histones H3 (Han et al., 2007; Weinberg et al., 2006) ce qui suggère que l'action du TGS pourrait se faire *via* la formation locale d'hétérochromatine.

CHAPITRE 3 – LE COUPLAGE EPISSAGE - CHROMATINE

L'idée que la structure de la chromatine puisse affecter l'épissage alternatif a été suggérée pour la première fois par l'équipe d'A.R. Kornblihtt en 2001. En utilisant des plasmides réplicatifs ou non contenant un mini-gène FN1, les auteurs montrent que l'inclusion de l'exon cassette EDI dans l'ARNm est augmentée en réponse à la réplication de la matrice d'ADN. Il a été ainsi suggéré que la structure plus compacte du plasmide une fois chromatinisé après réplication (Cereghini and Yaniv, 1984) agit comme une barrière et pourrait ralentir l'élongation de l'ARN polymérase II (Kadener et al., 2001).

1. LES EXONS ET LES INTRONS SONT ASSOCIES A DES MARQUES

CHROMATINIENNES DIFFERENTES

1.1. Les nucléosomes sont préférentiellement localisés au niveau des exons

L'analyse du positionnement des nucléosomes à grande échelle chez plusieurs eucaryotes supérieurs dont l'homme, la drosophile, *Caenorhabditis elegans*, montre que les nucléosomes sont plus fréquemment positionnés (approximativement 1,5 fois) dans les exons internes que dans les régions introniques flanquantes, c'est-à-dire de 300 à 1000 pb de part et d'autre des extrémités des exons (Andersson et al., 2009; Chodavarapu et al., 2010; Schwartz et al., 2009; Spies et al., 2009; Tilgner et al., 2009). **(Figure 22)** Il semblerait aussi que les introns dans leur globalité soient moins enrichis en nucléosomes comparés aux exons (Andersson et al., 2009).

Par ailleurs, il est intéressant de constater que la taille moyenne des exons et la longueur de la séquence d'ADN enroulée autour des nucléosomes sont toutes deux d'environ 150pb.

De plus, les données mettent aussi en évidence que cet enrichissement en nucléosomes est plus prononcé dans les cas des exons flanqués par des sites d'épissages faibles (Tilgner et al., 2009). Au contraire, les nucléosomes sont moins présents au niveau des pseudo-exons, des séquences introniques ayant les caractéristiques d'un exon (séquences consensus des sites d'épissage 5' et 3') mais qui ne sont jamais inclus dans l'ARNm (Tilgner et al., 2009). **(Figure 22)**

De ces corrélations, il a été proposé que les nucléosomes localisés au niveau des exons puissent faciliter leur reconnaissance parmi l'immensité des introns et les distinguer des pseudo-exons.

En outre, il ne faut pas écarter la possibilité que le lien entre le positionnement des nucléosomes et l'architecture des exons et des introns soit indirect et permis, par exemple, par une composition nucléotidique similaire des séquences préférentielles du positionnement du nucléosome et celles affines pour certains facteurs d'épissage.

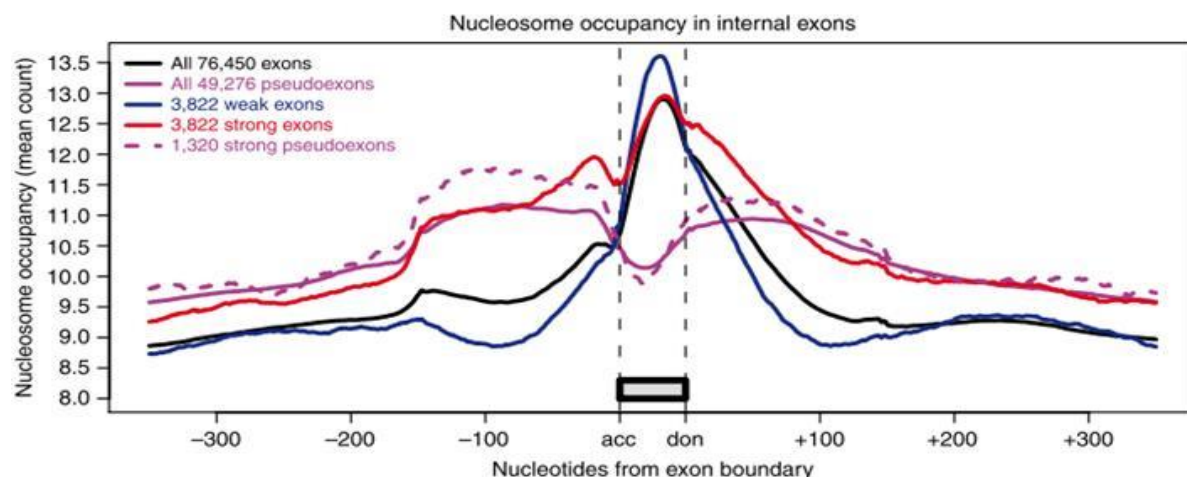


Figure 22. Profil du positionnement du nucléosome au niveau des exons constitutifs internes. Le profil est issu de l'analyse des données de ChIP-seq effectué dans des cellules humaines T CD4⁺. (Tilgner et al., 2009)

1.2. Les exons et les introns sont-ils enrichis en modifications épigénétiques d'histones différentes ?

Pour évaluer si les nucléosomes enrichis au niveau des exons sont de composition différentes, et donc potentiellement de fonctions différentes, plusieurs équipes ont examiné les distributions globales des modifications épigénétiques d'histones au niveau des jonctions exons/introns (Andersson et al., 2009; Dhami et al., 2010; Hon et al., 2009; Huff et al., 2010; Kolasinska-Zwierz et al., 2009; Schwartz et al., 2009; Spies et al., 2009).

Un intérêt particulier a été porté à la modification H3K36me₃, un marqueur de l'élongation transcriptionnelle (Bell et al., 2007). Cette marque est enrichie dans les exons comparés aux séquences introniques flanquantes (Andersson et al., 2009; Kolasinska-Zwierz et al., 2009; Schwartz et al., 2009; Spies et al., 2009; Tilgner et al., 2009) (**Figure 23A**). Cependant, bien que Spies *et al.* trouvent un enrichissement significatif de H3K36me₃ sur les exons (Spies et al., 2009), cela pourrait aussi être une conséquence de l'enrichissement en nucléosomes (Tilgner et al., 2009) (**Figure 23B**). Cette contradiction s'explique par la prise

en compte ou non des niveaux d'occupation des nucléosomes. Des conclusions différentes quant à l'enrichissement d'autres marques d'histones, comme pour H4K20me1, sont encore obtenues pour les mêmes raisons (Spies et al., 2009) (Hon et al., 2009).

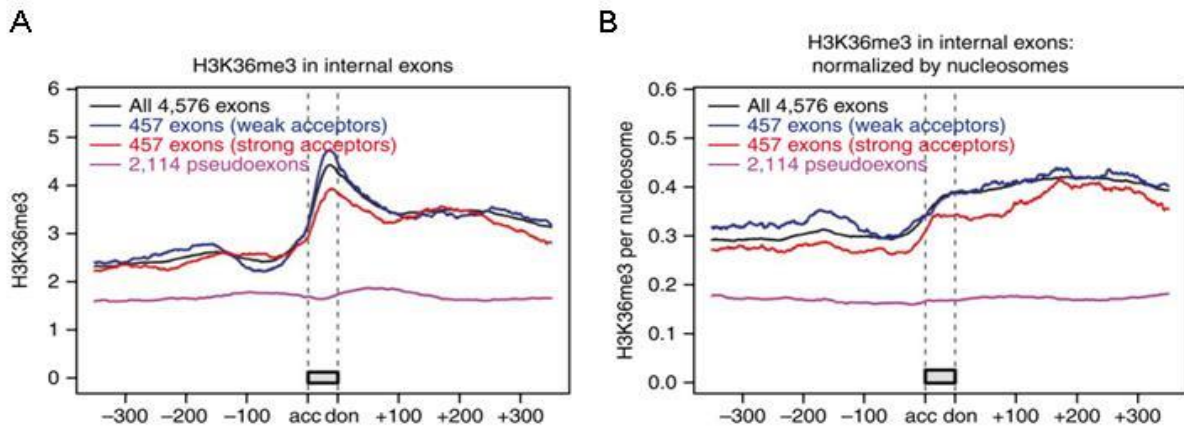


Figure 23. Profil de H3K36me3 dans les exons constitutifs internes. Le profil est issu de l'analyse des données de ChIP-seq effectué dans des cellules humaines T CD4+. Le signal de H3K36me3 n'a pas été (A) ou a été (B) normalisé par celui des nucléosomes. (Tilgner et al., 2009)

1.3. Les exons et les introns sont-ils enrichis en variants d'histones différents ?

Les nucléosomes peuvent aussi se différencier par le type de variants d'histones qu'ils contiennent. Ainsi, des variants d'histones spécifiques pourraient participer à la définition des exons.

Le variant d'histone H2A.Bbd est associé aux gènes transcriptionnellement actifs et ce uniquement pour les gènes multi-exoniques (c'est-à-dire contenant un ou plusieurs introns). De plus, ce variant est enrichi au niveau des sites d'épissage 3' et ce de manière plus marquée qu'avec le variant H2A.Z. Cependant, la comparaison avec H2A.Z n'est que peu informative puisque dans la même étude les auteurs ont caractérisé ce variant comme spécifique du site d'initiation de la transcription. (Tolstorukov et al., 2012)

2. LA CHROMATINE AFFECTE L'ÉPISSAGE ALTERNATIF

La chromatine et la transcription étant étroitement liées, la chromatine pourrait moduler l'épissage alternatif selon les deux modèles que nous avons présenté précédemment pour la régulation de l'épissage par la transcription (cf. paragraphe 4.1.2 du chapitre 1). Dans cette

partie, nous allons présenter des études qui appuient ces modèles.

2.1. Le modèle de recrutement : la chromatine influence le recrutement de la machinerie d'épissage

Dans ce modèle, il est proposé que des facteurs d'épissage puissent interagir de manière directe ou indirecte, *via* des « adaptateurs » à la chromatine.

2.1.1. Recrutement des facteurs d'épissage au promoteur

La marque H3K4me3 est une marque associée au site d'initiation de la transcription des gènes actifs. CHD1 est l'une des protéines qui interagit spécifiquement avec H3K4me3. CHD1 interagit avec des composants du spliceosome, notamment le complexe SF3A (Sims et al., 2007; Tai et al., 2003). Cependant, la protéine CHD1 affecte l'efficacité d'épissage aussi bien dans un système *in vitro* qu' *in vivo* (Sims et al., 2007). Par conséquent, le rôle de CHD1 sur la régulation de l'épissage ne nécessite pas obligatoirement sa liaison à la chromatine. Néanmoins, *in vivo* CHD1 pourrait agir sur l'épissage uniquement quand ce facteur est lié à la chromatine mais cet aspect n'a pas encore été exploré.

2.1.2. Recrutement des facteurs d'épissage à l'intérieur des gènes

Plusieurs études suggèrent que les modifications d'histones modulent l'épissage en favorisant le recrutement de facteurs d'épissage. La première d'entre elles a été la triméthylation de la lysine 36 de l'histone H3 (H3K36me3), une marque détectée à l'intérieur des gènes transcriptionnellement actifs (Bell et al., 2007). Le gène *FGFR2* (*Fibroblast Growth Factor Receptor*) possède deux exons mutuellement exclusifs IIIb et IIIc qui sont inclus dans les cellules épithéliales ou mésenchymateuses respectivement. En comparant deux types cellulaires (cellules épithéliales PNT2 *versus* cellules mésenchymales hMSC), R. Luco *et al.* ont corrélé l'inclusion de l'exon IIIc et avec un enrichissement de H3K36me3 sur les histones H3 (Luco et al., 2010). Basé sur les résultats d'expériences de déplétion par ARN interférence et de RNA-IP, les auteurs suggèrent que la protéine MRG15, qui interagit spécifiquement avec H3K36me3, recrute le facteur d'épissage répresseur PTB au locus de l'exon IIIb et empêche l'inclusion de ce dernier (Luco et al., 2010).

Aussi, la marque H3K36me3 est reconnue par l'isoforme courte d'épissage de la protéine Psip1/Ledgf, p52, (Pradeepa et al., 2012) qui elle-même interagit avec la protéines SR ASF/SF2 (Ge et al., 1998; Pradeepa et al., 2012). Des essais d'épissage *in vitro* ont montré que p52 et ASF/SF2 agissent en synergie sur l'épissage (Ge et al., 1998). De plus, *in vivo*, la déplétion de p52 affecte l'épissage alternatif de nombreux gènes, dont quelques exemples dépendent de ASF/SF2 (Pradeepa et al., 2012). L'alliance de ces deux études représente un argument en faveur d'une modulation de l'épissage par H3K36me3 *via* le recrutement de p52.

Les variants d'histones pourraient aussi influencer les décisions d'épissage *via* le recrutement de facteurs d'épissage. Chez l'homme, le variant H2A.Bbd, qui comme précédemment évoqué est enrichi dans les exons, affecte l'efficacité d'épissage puisque sa déplétion est corrélée avec une augmentation de séquences contenant des jonctions exon/intron (Tolstorukov et al., 2012). De plus, H2A.Bbd est associé à des protéines impliquées dans la maturation des ARN notamment le complexe spliceosomal SF3b ce qui amène les auteurs à proposer un modèle où H2A.Bbd serait impliqué dans la formation d'une structure chromatinienne qui facilite l'épissage (Tolstorukov et al., 2012). Cependant, ce variant d'histone est encore plus enrichi au niveau du site de terminaison de la transcription qu'au niveau des exons (Tolstorukov et al., 2012); or l'équipe de W. Keller a précédemment montré que le complexe SF3b interagit avec la protéine CPSF impliquée dans le clivage des ARNpré-m au site de terminaison de la transcription (Kyburz et al., 2006). Ainsi, l'association entre H2A.Bbd et le complexe SF3b pourrait aussi être permise par l'intermédiaire de CPSF au site de terminaison de la transcription plutôt qu'au niveau des exons. Cette possibilité n'a pas été évoquée.

2.2. Le modèle cinétique : la chromatine affecte la vitesse d'élongation transcriptionnelle

2.2.1. Les nucléosomes représentent une barrière pour l'ARN polymérase II

Pendant l'élongation transcriptionnelle, l'ARN polymérase II doit surmonter la barrière imposée par les nucléosomes. Plusieurs études *in vitro* ont montré que la présence d'un nucléosome est suffisante pour ralentir ou stopper la transcription par l'ARN polymérase II (Hodges et al., 2009; Izban and Luse, 1991) (**Figure 24A**). De plus, une approche *in vivo* basée sur le séquençage des extrémités 3' des transcrits suggère aussi que l'ARN polymérase

II pause au niveau des nucléosomes (Churchman and Weissman, 2011). Enfin, l'enrichissement relatif de l'ARN polymérase II et des nucléosomes au niveau des exons par rapport aux séquences introniques flanquantes (Chodavarapu et al., 2010) peut être interprété par un ralentissement ou une pause de l'ARN polymérase II au niveau des exons à cause de la présence des nucléosomes (**Figure 24B**).

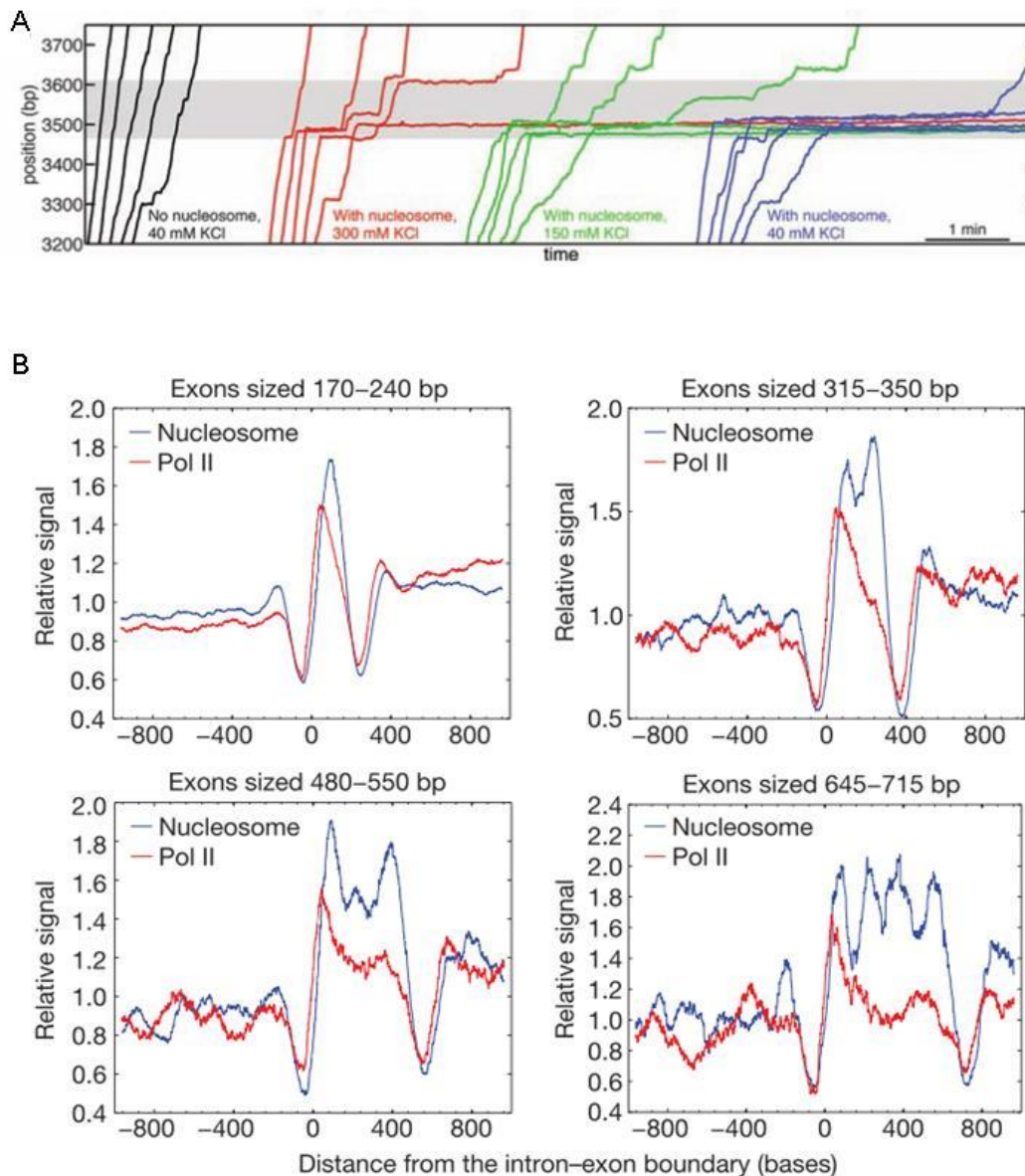


Figure 24. Les nucléosomes représentent une barrière pour l'ARN polymérase II *in vivo* et *in vitro*. **A.** Trajectoires représentatives de l'ARN polymérase II en cours de transcription sur des matrices d'ADN chromatinisées ou non et à différentes conditions salines. N=5 (Hodges et al., 2009). **B.** Signaux relatifs des nucléosomes et de l'ARN polymérase II à la jonction intron-exon. Les exons ont été groupés en fonction de leur taille permettant le positionnement de au plus 1, 2, 3 ou 4 nucléosomes (Chodavarapu et al., 2010).

2.2.2. Des marques chromatinienne et leurs facteurs affecteraient l'épissage en modulant la vitesse d'élongation de la polymérase II

a. Le complexe SWI/SNF

SWI/SNF est un complexe qui remodèle les nucléosomes, notamment au niveau des régions promotrices permettant le recrutement de facteurs de transcription. BRM, une des sous-unités catalytiques de ce complexe interagit avec les snRNA U1 et U5 et le facteur d'épissage Sam68 (Batsché et al., 2006) connu pour réguler l'épissage de CD44 (Matter et al., 2002). L'inclusion des exons alternatifs de plusieurs gènes dont CD44 est corrélée au niveau d'expression de BRM (Batsché et al., 2006). De plus, l'inclusion des exons alternatifs de CD44 coïncide avec un enrichissement de BRM et de l'ARN polymérase II, notamment sa forme phosphorylée en S5 (associée à une polymérase en pause), au locus où se trouvent les exons alternatifs (Batsché et al., 2006). Ceci suggère que BRM pourrait affecter les décisions d'épissage alternatif de CD44 *via* un ralentissement ou des pauses de l'ARN polymérase II. L'activité ATPase requise pour le remodelage de la chromatine n'est pas nécessaire au rôle de BRM dans la régulation de l'épissage alternatif de CD44. En revanche, son domaine d'interaction à la chromatine est nécessaire pour favoriser l'inclusion des exons variants. Enfin la déplétion de SNF5, une autre sous-unité du complexe SWI/SNF, affecte de la même manière l'épissage de CD44, suggérant que le rôle de BRM sur la modulation de l'épissage dépend de sa fonction au sein du complexe SWI/SNF (Batsché et al., 2006).

La proposition d'une implication directe de Brm dans la régulation de l'épissage alternatif a été renforcée par les images de microscopie électronique qui montre Brm associée aux pré-mRNP naissant dans des anneaux de Balbiani de *Chironomus tentans* (Tyagi et al., 2009).

La déplétion par ARN interférence de plusieurs sous-unités (Brm, Snr1 et Mor) du complexe SWI/SNF chez la drosophile affecte de manière similaire l'épissage alternatif de plusieurs gènes (Waldholm et al., 2011). Ainsi, Brm pourrait affecter l'épissage au sein du complexe SWI/SNF ; cependant les auteurs de ce papier n'ont fourni aucun argument en faveur d'un rôle direct (c'est-à-dire au locus des exons alternatif concernés) de ces protéines sur la régulation de l'épissage des gènes testés.

b. Les modifications d'histones

Les modifications d'histones sont associées à des états chromatinien différents. En postulant qu'à l'intérieur des gènes actifs, des marques d'histones différentes sont aussi

associées à des structures chromatinienne différentes, il a été proposé que ces modifications puissent moduler l'épissage en régulant la vitesse d'élongation de l'ARN polymérase II : une chromatine relâchée pourrait « accélérer » la polymérase alors qu'une chromatine compacte pourrait « freiner » la polymérase.

Acétylation des histones :

L'acétylation des histones est associée à un état chromatinien plus accessible (Loyola et al., 2001). I.E. Schor *et al.* ont mis en évidence que lors de la dépolarisation de neurones, le niveau d'inclusion de l'exon 18 de NCAM (*Neural Cell Adhesion Molecule*) est inversement corrélé au niveau de l'acétylation K9 de l'histone H3 au niveau de l'exon 18 (Schor et al., 2009). De plus, le mutant C4 de l'ARN polymérase II, qui est associé à une vitesse d'élongation plus lente, favorise l'inclusion de l'exon 18 de NCAM (Schor et al., 2009). Ainsi, les auteurs ont proposé un modèle dans lequel l'inclusion de l'exon 18 est défavorisée lorsque le locus de l'exon 18 est acétylé en raison de l'augmentation de l'élongation transcriptionnelle. Par la suite, d'autres études ont corrélé les niveaux d'acétylation des histones, l'enrichissement relatif de l'ARN polymérase II par ChIP interprété comme relatif à sa vitesse et l'inclusion d'exons alternatifs (Hnilicová et al., 2011; Khan et al., 2013).

Méthylation des histones :

La triméthylation de la lysine 9 de l'histone H3 (H3K9me3) et les protéines HP1 liées ont été initialement décrit comme associées à des régions transcriptionnellement inactives. Néanmoins, au laboratoire V. Saint-André *et al.* ont mis en évidence la présence de H3K9me3 à l'intérieur du gène CD44 et de manière intéressante un pic de cette modification d'histone est localisé dans le groupe d'exons variants. En utilisant plusieurs systèmes d'étude, les auteurs ont mis en évidence que les niveaux de la modification d'histone H3K9me3 et ceux de l'un de ses lecteurs, HP1 γ , sont corrélés au niveau d'inclusion des exons alternatifs de CD44 (Saint-André et al., 2011). La présence de HP1 γ au locus des exons variants de CD44 est aussi associée à un enrichissement de l'ARN polymérase II, du facteur associé à la snRNP U2, U2AF65 et de Prp8 (Saint-André et al., 2011). L'ensemble de ces résultats suggère que la protéine HP1 γ recrutée sur H3K9me3 pourrait affecter les décisions d'épissage de CD44 *via* un ralentissement ou des pauses de l'ARN polymérase II.

Par ailleurs, une étude à l'échelle du génome a mis en évidence que HP1 γ est présent dans le corps des gènes transcriptionnellement actifs et sa déplétion est associée à une diminution de l'efficacité d'épissage (Smallwood et al., 2012).

Ubiquitination des histones :

USP49 est une dé-ubiquitinase spécifique de l'histone H2B (Zhang et al., 2013). La déplétion par ARN interférence d'USP49 affecte l'épissage alternatif de nombreux gènes. De plus, les auteurs suggèrent qu'USP49 agit sur l'épissage co-transcriptionnel puisque sa déplétion réduit la localisation des snRNP U1 et U2 sur la chromatine et diminue l'efficacité d'excision de l'intron 2 de SLMO2 enrichi en ubiquitination (Zhang et al., 2013).

c. Les protéines Argonaute

Comme nous l'avons présenté dans le chapitre 2 de cette introduction, plusieurs études suggèrent que des petits ARN ciblant des régions promotrices peuvent induire la formation d'hétérochromatine localement *via* les protéines Argonaute (AGO).

L'équipe d'A.R. Kornblihtt s'est interrogée sur un rôle potentiel sur l'épissage des petits ARN ciblant le corps des gènes. M. Allo *et al.* montrent que la transfection de petits ARN interférants (siRNA) ciblant la région intronique ou exonique proche de l'exon cassette EDI de la fibronectine favorise l'inclusion de cet exon. Cet effet qui est accompagné d'un enrichissement local en marques répressives H3K9me2 et H3K27me3 est inhibé lors de la déplétion des protéines Argonaute (AGO) 1 et 2. Les auteurs suggèrent que les siRNA pourraient induire une hétérochromatinisation locale *via* AGO ce qui pourrait ralentir la polymérase et favoriser l'inclusion de l'exon cassette EDI. (Alló et al., 2009)

De plus, au laboratoire, M. Ameyar *et al.* ont mis en évidence que les protéines AGO1 et AGO2 sont associées à des facteurs de la chromatine mais aussi à des composants du spliceosome et des facteurs d'épissage (Ameyar-Zazoua et al., 2012). La déplétion d'AGO1 et AGO2 induit une diminution de l'inclusion des exons variants de CD44 et est corrélée à une diminution de la présence de l'ARN polymérase II et des composants du spliceosome Prp6 et Prp8. Les auteurs suggèrent qu'AGO1 et AGO2 favorisent le ralentissement de la polymérase et le recrutement du spliceosome localement pour faciliter l'inclusion des exons variants de CD44. Les protéines AGO pourraient être un élément initiateur de l'action de HP1 γ en

réponse au PMA puisque leur déplétion empêche aussi l'enrichissement en H3K9me3 (Saint-André et al., 2011).

2.3. Bilan de l'impact de la chromatine sur l'épissage alternatif

2.3.1. Les facteurs de la chromatine peuvent réguler l'épissage indépendamment de leurs fonctions chromatinienne

Plusieurs études présentent des arguments en faveur d'un rôle des facteurs de la chromatine sur la régulation de l'épissage. Mais il est légitime de se demander si de tels facteurs modulent l'épissage grâce à leurs fonctions chromatinienne connues ou s'ils mettent en jeu une nouvelle fonction. En effet, l'action de BRM sur l'épissage de CD44 est indépendante de son domaine ATPase octroyant l'activité de remodelage de la chromatine. En revanche, son bromodomaine permettant la liaison aux queues d'histones acétylées est indispensable (Batsché et al., 2006). Ces données suggèrent que BRM, associé à la chromatine, a des fonctions différentes qui mettent en jeu des domaines protéiques différents. De la même façon, l'histone-arginine méthyltransférase CARM1 peut affecter l'épissage alternatif sans recourir à son activité enzymatique (Ohkura et al., 2005).

2.3.2. Le modèle de recrutement et le modèle cinétique ne sont pas nécessairement exclusifs

Le modèle de recrutement et le modèle cinétique sont conciliables et plusieurs études le suggèrent. Par exemple, la sous-unité BRM du complexe SWI/SNF semble appuyer l'hypothèse des modèles coopératifs. En effet, comme exposé précédemment BRM interagit avec des composants du spliceosome et le facteur d'épissage Sam68 ; mais son action sur l'épissage des exons alternatifs de CD44 semble aussi liée à un ralentissement de l'ARN polymérase II (Batsché et al., 2006).

Les protéines de liaison à l'ARN, Hu, apportent aussi des arguments en faveur d'un modèle bi-causal. Les protéines Hu défavorisent l'inclusion des exons alternatifs 23a et 6 de Nf1 et Fas respectivement (Zhou et al., 2011). Les auteurs suggèrent que les protéines Hu, qui interagissent et inhibent HDAC2 (=recrutement), induisent une hyperacétylation des histones localement. Ceci a pour conséquence de provoquer l'augmentation de la vitesse d'élongation transcriptionnelle (=cinétique) et l'exclusion des exons alternatifs (Zhou et al., 2011).

2.3.3. Mais au final, quel est le modèle ?

De nombreux modèles de couplage épissage - chromatine ont été apportés suggérant qu'il n'y a pas un modèle unique de couplage. Cette observation est cohérente avec le caractère hautement modulable de la machinerie d'épissage (cf. chapitre 1). En effet, si la chromatine influence localement l'épissage, elle n'en est pas un régulateur majeur : les régulateurs majeurs sont le spliceosome et les facteurs d'épissage associés. Donc, si la chromatine régule l'épissage localement, son effet est dépend d'un contexte particulier : c'est-à-dire d'un contexte où le choix d'un site d'épissage plutôt qu'un autre peut basculer facilement, laissant place à d'autres paramètres pour réguler l'épissage. En fonction du gène regardé, ou de l'exon observé, des sites d'épissage et des séquences *cis* régulatrices impliquées mais aussi en fonction du type cellulaire utilisé où fluctue l'expression des différents facteurs d'épissage... une même marque chromatinienne ou un même facteur chromatinien pourrait réguler ou non une décision d'épissage. Pour illustrer ce point, la déplétion de HP1 γ dans les cellules HeLa affecte l'épissage de CD44 (Saint-André et al., 2011) alors qu'elle n'a pas d'effet dans les cellules HCT116 (Smallwood et al., 2012).

En ce sens, trouver une marque chromatinienne « universelle » associée aux exons alternatifs semble improbable. Des critiques récurrentes du modèle du couplage épissage - chromatine viennent du fait que les changements d'épissage sont parfois associés à des changements chromatiniens non localisés sur l'exon alternatif mais sur des régions plus vastes, parfois sur tout le locus génique (Hnilicová et al., 2011; Luco et al., 2010). De mon point de vue, cette critique ne semble que peu justifiée puisque des changements chromatiniens sont susceptibles d'affecter uniquement l'épissage des exons alternatifs mais pas ceux des exons constitutifs.

Au final, dans un contexte précis la chromatine et notamment les nucléosomes enrichis au niveau des exons pourraient momentanément ralentir ou pauser l'ARN polymérase II et favoriser les échanges moléculaires entre l'ARN polymérase II, les facteurs d'épissage, la chromatine et l'ARNpré-m.

Les études du couplage épissage - chromatine ont principalement porté sur l'impact des modifications d'histones. Les modifications épigénétiques ont aussi lieu sur l'ADN. Plusieurs études ont mis en évidence un enrichissement de la méthylation de l'ADN sur les exons

comparés aux introns (Chodavarapu et al., 2010; Feng et al., 2010; Ficiz et al., 2011; Laurent et al., 2010). Mon premier projet de thèse a consisté à étudier l'impact potentiel de la méthylation sur la régulation de l'épissage.

PROJET 1 : L'IMPACT DE LA METHYLATION DE L'ADN SUR L'EPISSAGE ALTERNATIF

INTRODUCTION

Etat des connaissances du sujet au début de ma thèse

Les niveaux de méthylation aux régions promotrices et leurs implications dans l'expression génique ont été intensément étudiés. Pourtant d'autres régions génomiques sont affectées par la méthylation de l'ADN, notamment l'intérieur des gènes. Bien que le corps des gènes soit pauvre en dinucléotides CpG, il est riche en méthylation de l'ADN (Laurent et al., 2010; Lister et al., 2009). Par ailleurs, même si le corps des gènes est hautement méthylé, le gène peut être transcrit. Et même, il semble qu'il y ait une corrélation positive entre la méthylation intragénique et les niveaux d'expression des gènes (Ball et al., 2009; Hellman and Chess, 2007; Lister et al., 2009; Rauch et al., 2008). Cela mène à un apparent paradoxe où la méthylation au niveau des promoteurs est inversement corrélée à l'expression alors que la méthylation intragénique est corrélée positivement à l'expression. La méthylation de l'ADN est donc associée à une inhibition de l'initiation transcriptionnelle mais pas à celle de l'élongation transcriptionnelle.

Quel est donc le rôle de cette méthylation intragénique ? Plusieurs hypothèses ont été documentées. D'une part, la méthylation intragénique pourrait permettre l'inhibition des SINEs (*short interspersed nuclear elements*), des LINEs (*long interspersed nuclear elements*) et des rétrovirus endogènes qui sont hautement méthylés (Yoder et al., 1997). D'autre part, la méthylation des corps des gènes pourrait inhiber l'initiation transcriptionnelle à partir de promoteurs « orphelins », c'est-à-dire des promoteurs non annotés (Illingworth et al., 2010). En accord avec cette idée, une étude suggère que la méthylation de l'ADN régule les choix des promoteurs alternatifs (Maunakea et al., 2010). Néanmoins, ces hypothèses concernent uniquement l'implication potentielle de la méthylation de l'ADN à des sites intragéniques précis. Pourtant la méthylation de l'ADN est répartie tout au long du corps des gènes ce qui suggère que la méthylation de l'ADN pourrait être impliquée dans d'autres fonctions.

L'étude du méthylome par séquençage à haut débit après traitement au bisulfite de sodium a permis de mettre en évidence que les exons sont enrichis en dinucléotides CpG méthylées comparés aux introns (Chodavarapu et al., 2010; Choi, 2010; Laurent et al., 2010). La transition du degré de méthylation se fait au niveau de la jonction exon-intron (Laurent et

al., 2010) suggérant un rôle de la méthylation intragénique dans la régulation de l'épissage. De plus, la comparaison des méthylomes et des transcriptomes des abeilles reines et travailleuses ont mis en évidence que certaines différences d'épissage alternatif sont corrélées avec des différences de méthylation de l'ADN (Lyko et al., 2010). Les mécanismes possibles qui modulent l'épissage alternatif par la méthylation de l'ADN n'ont pas été explorés dans cette étude. De manière intéressante, l'efficacité d'élongation transcriptionnelle d'un mini-gène intégré dans le génome est diminuée lorsque le corps de ce mini-gène est méthylé (Lorincz et al., 2004), ainsi, cette marque épigénétique pourrait affecter l'épissage selon un modèle cinétique. Par ailleurs, bien que son action soit indépendante de sa capacité de liaison à la méthylation de l'ADN, MeCP2 affecte l'épissage alternatif *via* son interaction avec le facteur d'épissage YB-1 (Young et al., 2005). La méthylation de l'ADN pourrait donc réguler l'épissage alternatif *via* les protéines de liaison à l'ADN méthylé en recrutant des facteurs d'épissage. Aussi, par analogie à la coopération établie entre DNMT1 et les protéines HP1 au niveau des promoteurs pour réprimer l'expression génique (Smallwood et al., 2007), ces protéines pourraient s'associer dans le corps des gènes pour réguler l'épissage.

Dans cette étude, nous avons montré que la teneur en dinucléotides CpG est plus importante dans les exons que dans les introns ce qui corrèle avec le différentiel des niveaux de méthylation de l'ADN. En accord avec les données méta-génomiques sur les CpG et les niveaux de méthylation de l'ADN, les exons des CD44 sont enrichis en dinucléotides CpG et en méthylation de l'ADN comparés aux introns. En revanche, il n'y a pas de différences significatives de méthylation entre les exons constants et les exons alternatifs. De plus, les profils de méthylation CD44 ne sont pas corrélés aux décisions d'épissage. Néanmoins, les enzymes DNMT affectent l'épissage de nombreux gènes et une approche à l'échelle du génome suggère que les exons régulés par les enzymes DNMT sont plus enrichis en sites méthylés. Ensemble, nos résultats suggèrent que l'épissage est impacté par la méthylation de l'ADN, néanmoins le mécanisme mis en jeu reste encore à définir.

MATERIELS ET METHODES

Analyse bioinformatique

Les coordonnées des transcrits et de leurs exons (RefSeq Genes) du génome de souris (assemblage mm9) ont été extraites à partir de la plate-forme internet Galaxy (<https://usegalaxy.org/>). Pour chaque transcrit, les séquences des premiers, internes et derniers exons et introns ont été extraites *via* Galaxy. Les séquences des 5000 pb, 2000 pb et 1000 pb en amont et en aval de chaque transcrit ont aussi été extraites. Dans le cas des transcrits alternatifs provenant d'un même gène, les mêmes éléments retrouvés dans des catégories différentes ont été comptabilisés dans chaque catégorie (par exemple, si selon les transcrits un exon peut être interne ou terminal, il est pris en compte dans les deux catégories). L'analyse des proportions en nucléotides a été faite sur python (v2.7.2). L'analyse des données transcriptomiques issues à partir de puces exons et de séquençage à haut débit ainsi que l'analyse des données méthylomiques ont été effectuées par Genosplice (<http://www.genosplice.com>). Les séquences des transcrits du génome humain (hg19) ont été extraites *via* Galaxy. L'analyse de la présence des régions *seed* des siRNA a été faite sur python (v2.7.2).

Culture cellulaire, traitements, et transfection de petits ARN interférents

Les cellules MCF7 (ATCC, HTB-22), SKOV3 (ATCC, HTB-77), SW626 (ATCC, HTB-78) ont été cultivées dans du milieu *Dulbecco's modified Eagle* (Gibco, #31966-021) contenant 10% (v/v) de sérum fœtal de bovin (Thermo Scientific, #SV30160.03) et 100 U/mL de pénicilline-streptomycine (Gibco, #15140-122). Les cellules HeLa (ATCC, CCL-2) ont été cultivées dans le même milieu complémenté de 7% (v/v) de sérum fœtal de bovin. Les cellules HeLa ont été traitées avec 40 ng/mL de PMA (phorbol-12-myristate-13-acetate) pendant 4 h avant de lyser les cellules. Les cellules HeLa ont été traitées avec 1 μ M d'AZAdC (5-aza-2'-deoxycytidine) toutes les 12 h pendant 3 jours avant de lyser les cellules. Les cellules HeLa ont été transfectées selon les instructions du fabricant avec un mélange contenant des petits ARN interférents (siRNA) (20 nM) et l'agent transfectant RNAi Max (Life Technologies, #13778075). Pour MeCP2, MBD1 à 4, les cellules ont été transfectées une fois avec un mélange de 3 siRNA et lyser 3 jours après. Pour DNMT1, les cellules ont été transfectées deux fois, à 3 jours d'intervalle puis lyser 5 jours après la première transfection.

Extraction d'ARN, transcription inverse et PCR quantitative

Les cellules ont été lysées dans du TSM300 (10 mM de Tris pH 7,5, 300 mM de NaCl, 2 mM de MgCl₂, 1% de NP-40) sur la glace pendant 25 minutes. Après centrifugation, les surnageants contenant les ARN totaux ont été extraits sur des colonnes « *XS spin columns* » (Macherey Nagel, #740900.50) selon les instructions du fabricant. Après traitement à la DNase I (Roche, #04716728001), la qualité des ARN a été contrôlée sur gel d'agarose. Les bibliothèques d'ADNc ont été produites avec 0,5 µg d'ARN inversement transcrit avec la transcriptase inverse M-MLV (Life Technologies, #28025013) et des hexamères aléatoires (Sigma). Les PCR quantitatives en temps réel (qPCR) ont été effectuées dans des réactions de 10 µL avec le Brilliant III Ultra Fast SYBR-Green QPCR Master Mix (Agilent, #600882). Les réactions de PCR ont été réalisées sur le système Stratagene MX3005p selon le profil thermique suivant : 5 min à 95°C, puis 37 cycles de 10 secs à 95°C et 12 secs à 62°C. Les données de qPCR ont été analysées avec le logiciel MxPro comme décrit précédemment (Batsché et al., 2006).

Extraction d'ADN, immunoprécipitation de l'ADN méthylé

Les cellules ont été lysées pendant 5 h à 55°C dans un tampon (20 mM Tris-HCl, pH8, 4 mM EDTA, 20 mM NaCl, 1 % SDS) additionné de 700 µg/mL de protéinase K. L'ADN a ensuite été extrait par phénol/chloroforme, précipité à l'isopropanol et resuspendu dans du TE. Après un traitement à la RNase pendant 1 h à 37°C, l'ADN a été soniqué (Bioruptor) de manière à obtenir des fragments d'environ 300 pb. Dans le tampon IP (10 mM NaH₂PO₄, 10 mM Na₂HPO₄, 140 mM NaCl et 0,05% Triton X-100), 2 µg d'ADN soniqués ont ensuite été immunoprécipités avec 1 µg d'anticorps anti- 5-méthylcytosine (Diagenode, 33D3) pendant 4 h puis 2 h supplémentaires après l'addition de 40 µL de billes Dynabeads anti-souris (Life Technologies, #112.02D). Après 3 lavages dans le tampon IP et 2 lavages dans le tampon TE, l'ADN a été isolé dans 10% de chelex (Biorad, #142-1253).

Western Blot

Des extraits protéiques totaux ont été séparés par électrophorèse sur gel à gradient 4-12% d'acrylamide (Bio-Rad, #345-0124) et transférés sur une membrane de nitrocellulose (Bio-Rad, #1620115). Les anticorps suivants ont été utilisés : anti-DNMT1 (Active Motif, #39905), anti-BRG1 (Euromedex, #2SN-2E12-AS), anti-H3 (Abcam, #ab1791), anti-IgG HRP de lapin (GE Healthcare, #NA934V), anti- IgG HRP de souris (Abcam, #ab6808).

Liste des séquences des amorces et des siRNA utilisés

	Nom	Localisation	Séquence (5'-3')
amorces	CD44-PPF	PP	TGGCTGGGTGCATTTCTCTCAA
	CD44-PPR	PP	TGAAAGGAACCATGCCAGGACT
	CD44-i1F	i1	TGCAGCACATGGCAGAAAGTAACC
	CD44-i1R	i1	ACAGGAGCCTGTTAGTCCAGACA
	CD44-C2F	C2	TGCCGCTTTGCAGGTGTATT
	CD44-C2R	C2	GGCAAGGTGCTATTGAAAGCCT
	CD44-i2F	i2	ACAACTGCCACTGGCCAATCAT
	CD44-i2R	i2	TAGGCAGGCTGTGTGTGCAAAT
	CD44-C3F	i2-C3	ACAGGTATGGGTTCATAGAAGGGCAC
	CD44-C3R	C3	TGTGTCATACTGGGAGGTGTTGGA
	CD44-i3F	i3	AGAGGGTGCCTTTCTATCTGGCAA
	CD44-i3R	i3	TGCTTGGAAGCTAGACCTGGTGA
	CD44-C4F	C4	ACATCAGTCACAGACCTGCCCAAT
	CD44-C4R	i4	AACACACCTGAGCCCTTTCCC
	CD44-C5F	C5	AGCAGCACTTCAGGAGGTTACA
	CD44-C5R	C5	TGATCCAGGGACTGTCTTCGT
	CD44-i5F	i5	CCAAGGCAAACAACCTGGTGGAAA
	CD44-i5R	i5	TGCAATGCATGAGGGAGGCTTT
	CD44-v2F	v2	CAGCAACTGAGACAGCAACCAA
	CD44-v2R	v2	AACCAATCCCAGGTTTCTTGCC
	CD44-i7F	i7	GCAGTGTAGCTTCCATCTCTGTTTGG
	CD44-i7R	i7	TGGAAAGGTTTCACACGTGCCT
	CD44-v3F	v3	GGCTGGGAGCCAAATGAAGAAA
	CD44-v3R	v3	CATCATCATCAATGCCTGATCCAGA
	CD44-i8F	i8	AGGTGCATCTCTTATTCCAGCCCT
	CD44-i8R	i8	TGAGCAAGCCATTTCATAGCCAGTG
	CD44-v4F	v4	CAGTGGAACCCAAGCCATTCAA
	CD44-v4R	v4	CCTTGTGGTTGTCTGAAGTAGCAC
	CD44-i9F	i9	TGGTTTGTGGTTCTGCCTTTGACA
	CD44-i9R	i9	AGTGTTACCTCTGAGCATCCACTCTC
	CD44-v5F	v5	GAAACTGGAACCCAGAAGCACA
	CD44-v5R	v5	TGATGCTCATGGTGAATGAGGG
	CD44-i10F	i10	TAGGTGCCATGTGGACATCACCAA
	CD44-i10R	i10	ACACACTGGAAATGCACCAGCA
	CD44-v6F	v6	CAGAAGGAACAGTGGTTTGGCA
	CD44-v6R	v6	GTCTTCTTTGGGTGTTTGGCGA
	CD44-i11F	i11	TGGCTCAGAATGAAGCAAGGCA
	CD44-i11R	i11	GGCTGGAAGTGATCTCAGGAAAGGAA
	CD44-v7F	v7	TGCAAGGAAGGACAACACCAAG
	CD44-v7R	v7	GGGTGTGAGATTGGGTTGAAGA
	CD44-i12F	i12	TCTGGGATAACAGGGTCACCACAT
	CD44-i12R	i12	TGCACTGCAGCCTTTCAACACT
	CD44-v8F	v8	ACGCTTCAGCCTACTGCAAA
	CD44-v8R	v8	AAGAGGTCCTGTCCTGTCCAAA
	CD44-i13F	i13	TTGCACCTGCAACTCAGGTCAGTA
	CD44-i13R	i13	ATCACTGCACCAAGGAAACGCT
	CD44-v9F	v9	GAGCTTCTCTACATCACATGAAGGC

siRNA	CD44-v9R	v9	GTCAGAGTAGAAGTTGTTGGATGGTC
	CD44-i14F	i14	TTGCGCTGTGGTCCAATAGACTGT
	CD44-i14R	i14	ACCATGGAGGTCCAGAGGACAAA
	CD44-v10F	v10	ACCTCTCATTACCCACACACGA
	CD44-v10R	v10	TAGCTGAGGTCACTGGGATGAA
	CD44-i15F	i15	TGCCACATGCTTGGCCTCATT
	CD44-i15R	i15	TTCTTGAAGCTCTTGGGTAGGCCA
	CD44-i16F	i16	AGACATCTAGGGCACAAGGCAAGA
	CD44-i16R	i16	TGCGTGCAGCAAGATGAAACCT
	CD44-C18F	i17	TGTCTCTGAAGCTCACGCATGTCA
	CD44-C18R	C18	CGACTGTTGACTGCAATGCAAACCTGC
	CD44-i18F	i18	AGGGAAATGCAGGCCAACCAAA
	CD44-i18R	i18	ATGGGTGGTTGAGGGATGGAATGA
	CD44-C19F	C19	TGATCAACAGTGGCAATGGAGC
	CD44-C19R	C19	TCTGACGACTCCTTGTTCACCA
	MECP2-e3F		TGGACACGGAAGCTTAAGCAAAGG
	MECP2-e4R		TTCTTAGGTGGTTTCTGCTCTCGC
	MBD1-e13F		ATGATTCTGCCTCCAAATTGGCCC
	MBD1-e14R		CTTGGCAACCAGACTGCTGTATCT
	MBD2-e3F		CCCACAACGAATGAATGAACAGCC
	MBD2-e4R		AAGTCCTTGTAGCCTCTTCTCCCA
	MBD3-e1F		TTTACTATAGCCCGAGCGGGAAGA
	MBD3-e2R		GCGGCTCTTGTTTCATCTTGCTCAT
	MBD4-e5F		AAGAAATGGACACCTCCTCGGTCA
	MBD4-e6R		ACATCTCTCCAGTCTGCGGTTCTT
	H19-e2F		AGTGGACTTGGTGACGCTGTAT
	H19-e2R		TGGATGCTGTACTGTCTGCCAA
	DNMT1_6		CCCAATGAGACTGACATCAAA
	DNMT1_7		CGGGAAGTGAATGGACGTCTA
	DNMT1_1		AACGGTGCTCATGCTTACAAC
	MECP2_7		ACGGAGCGGATTGCAAAGCAA
	MECP2_8		CACCCAGGTCATGGTGATCAA
	MECP2_12		AGGCTGGACACGGAAGCTTAA
	MBD1_8		ACCGGGAACAGAGAATGTTTA
	MBD1_9		CGCGAAGTCTTTCGCAAGTCA
	MBD1_4		CAGGATCCGAAGCAAAGTTGA
	MBD2_8		AAGATGATGCCTAGTAAATTA
	MBD2_10		CGAAACGATCCTCTCAATCAA
	MBD2_7		TGGAAAGATGATGCCTAGTAA
	MBD3_5		GCCGGTGACCAAGATTACCAA
	MBD3_2		CCCGGAGATGGAGCACGTCTA
	MBD3_4		CGGGAAGAAGTTCCGCAGCAA
	MBD4_5		AAGCTTCTCATCGCTACTATA
	MBD4_6		CACGACGTAAAGCCTTTAAGA
	MBD4_3		CCGCCGAATGACCTCCGCAA

RESULTATS

1. META-ANALYSE DES PROPORTIONS DE NUCLEOTIDES DANS LES DIFFERENTS ELEMENTS D'UN GENE

Chez les mammifères, la méthylation de l'ADN est majoritairement localisée sur des cytosines des dinucléotides CpG. Nous avons effectué une étude bioinformatique pour déterminer s'il existe des biais nucléotidiques, notamment des biais de dinucléotides CpG, dans les exons constants *versus* les exons alternatifs. Pour cela nous avons analysé les séquences des différents éléments d'un méta-gène de souris (premier exon, exon interne, dernier exon, premier intron, intron interne et dernier intron) et de ses régions environnantes (les intervalles de 1 à 1000 pb ; de 1001 à 2000 pb et de 2001 à 5000 pb en amont et en aval des sites d'initiation et de terminaison transcriptionnelle respectivement) en distinguant les gènes multi-exoniques des mono-exoniques.

Dans un premier temps, nous avons analysé les proportions de chaque nucléotide A, T, C et G (**Figure 1A et 1B**). Il apparaît que les nucléotides ne sont pas répartis de manière équivalente et homogène dans les différents éléments du méta-gène. Les régions en amont et en aval des unités transcriptionnelles sont plus riches en A et T qu'en C et G. Dans le cas des gènes multi-exoniques, les premiers exons sont enrichis en C et G au dépend de leur composition en A et T ; ce qui est en accord avec le fait que plus de 70% des régions promotrices sont riches en dinucléotides CpG (Saxonov et al., 2006). Les exons internes sont majoritairement constitués de purines et sont plus riches en C et G que les introns. Dans le cas des gènes mono-exoniques, les exons présentent aussi un enrichissement en C et G comparé aux régions flanquantes, au dépend des A ; le niveau de T reste relativement constant.

Dans un second temps, nous avons analysé les proportions des dinucléotides CpG (**Figure 1C**). Comme attendu (Bird, 1980), la fréquence observée des dinucléotides CpG, mais pas celle des GpC est inférieure à la fréquence attendue en tenant compte de la proportion en C et en G dans chaque élément des unités transcriptionnelles. Par ailleurs, pour les gènes multi-exoniques, la haute fréquence de CpG dans le premier exon atteint quasiment la fréquence attendue. De manière intéressante, les exons internes sont enrichis en CpG. Mais cet enrichissement pourrait être une conséquence des proportions de GC (**Figure 1B et 1C**).

Ensuite, nous avons comparé les proportions des dinucléotides CpG dans l'ensemble des exons internes *versus* les exons définis comme des exons cassettes dans *Genome Browser*. Il apparaît que les exons cassettes sont composés de moins de CpG que l'ensemble des exons internes (**Figure 2**).

L'ensemble de ces résultats indique le pourcentage de GC ou le nombre de CpG pourrait être un marqueur des exons et la faible teneur en CpG pourrait être une propriété des exons cassettes. Ainsi, l'enrichissement en méthylation de l'ADN dans les exons *versus* les introns (Chodavarapu et al., 2010; Choi, 2010; Laurent et al., 2010) pourrait être une conséquence des proportions en CpG. En effet, en normalisant pas le nombre CpG, Lister *et al.* ne voient pas d'enrichissement en CpG méthylés dans les exons comparés aux introns (Lister et al., 2009).

2. ANALYSE DES NIVEAUX DE LA METHYLATION DE L'ADN DANS LE GENE CD44

Nous avons souhaité déterminer si le niveau de méthylation de l'ADN pourrait distinguer les exons constants des exons alternatifs.

Pour analyser les niveaux de méthylation de l'ADN nous avons effectué des expériences d'immunoprécipitation de l'ADN méthylé : MeDIP (Methylated DNA ImmunoPrecipitation). Après avoir extrait et soniqué l'ADN des cellules, cette expérience consiste à immunoprécipiter l'ADN méthylé en utilisant un anticorps anti 5-méthyl-cytosine. L'enrichissement relatif de méthylation de l'ADN a été analysé par PCR quantitative. (**Figure 3**)

Pour notre étude, nous avons choisi le gène CD44, un gène modèle de l'épissage alternatif de CD44. Chez l'homme, ce gène est composé d'un groupe de 9 exons alternatifs (ou variants), nommés v2 à v10, encadrés par 2 groupes d'exons constants, nommés C1 à C5 et C16 à C19. Dans les cellules HeLa, l'expression des exons variants représente au plus 5% de l'expression des exons constants (**Figure 4A**).

Nous avons mis en évidence que dans les cellules HeLa, les exons du gène CD44 sont plus méthylés que les introns (**Figure 4B**) ce qui est cohérent avec les résultats déjà publiés à

l'échelle du génome (Chodavarapu et al., 2010). En accord avec le fait que CD44 est un gène transcriptionnellement actif, la région promotrice proximale présente un niveau faible de méthylation de l'ADN.

Pour déterminer si les niveaux de méthylation de l'ADN sont proportionnels au nombre de CpG, nous avons calculé le nombre de ces dinucléotides dans les différents exons/introns du gène CD44. Précisément, nous avons calculé le nombre de CpG dans les régions correspondant à l'amplicon de qPCR et les 200 pb flanquants pour être cohérent avec la taille des fragments d'ADN soniqués lors du MeDIP. La région promotrice proximale et l'intron 1 (à l'extrémité 5') sont riches en CpG, ce qui est cohérent avec le fait que le gène CD44 est sous contrôle d'un promoteur à îlots CpG (**Figure 4C**). A l'exception de cette région promotrice, nous observons une corrélation entre les niveaux de méthylation de l'ADN et le nombre de CpG (**Figure 4C**), suggérant que les différents degrés de méthylation sont une conséquence du nombre de nucléotides méthylables. De plus, bien que la différence ne soit pas significative, le niveau de méthylation moyen des exons variants est inférieur à celui des exons constants (**Figure 4D**). Cette différence semble liée à la proportion de CpG (**Figure 4E**). Par ailleurs, nous soulignons que la proportion plus faible des CpG dans les exons variants *versus* constants dans le gène CD44 est cohérente avec les résultats précédents à l'échelle du génome.

3. DES PROFILS D'ÉPISSAGE DIFFÉRENTS NE SONT PAS CORRÉLÉS À DES NIVEAUX DE METHYLATIONS D'ADN DIFFÉRENTS

Le niveau basal de méthylation intragénique de CD44 semble distinguer, légèrement, les exons constants des exons variants. Nous avons alors souhaité déterminer si des profils d'épissage différents sont associés à des profils de méthylation de l'ADN différents.

3.1. Le traitement au PMA n'affecte pas la méthylation de l'ADN à l'intérieur du gène CD44

Dans un premier temps, nous avons traité des cellules HeLa au PMA (phorbol-12-myristate-13-acetate), une molécule connue pour induire l'inclusion des exons variants de CD44 (Matter et al., 2002). En effet, après 4 h de traitement, les exons variants de CD44 sont

4 à 5 fois plus exprimés dans l'ARNm alors que l'expression des exons constants est inchangée (**Figure 5A**). L'analyse par MeDIP ne révèle aucun changement de méthylation de l'ADN à la suite du traitement au PMA (**Figure 5B**).

Ainsi, la méthylation de l'ADN ne semble pas intervenir lors de la régulation de l'épissage alternatif induite par le traitement au PMA. Cependant, la méthylation de l'ADN étant définie comme une marque épigénétique stable et associée à des régulations sur le long terme, la durée limitée du traitement pourrait être un facteur limitant à l'action de la méthylation de l'ADN. Nous avons alors utilisé un autre système d'étude : des lignées cellulaires avec des profils d'épissage de CD44 différents.

3.2. Des cellules différentes avec différents profils d'épissage alternatif de CD44 ont des profils de méthylation similaires

Nous avons utilisé 3 lignées cellulaires différentes :

- les cellules MCF7 : cellules épithéliales provenant d'un adénocarcinome mammaire,
- les cellules SKOV3 : cellules de morphologie épithéliale provenant d'un adénocarcinome ovarien,
- les cellules SW626 : cellules de morphologie épithéliale provenant d'un adénocarcinome ovarien.

Dans les cellules MCF7, l'expression des exons variants v2 à v7 représente moins de 10% de celle des exons constants et l'expression des exons v8 à v10 représente entre 40 et 80% de celle des exons constants (**Figure 6A**). Dans les cellules SKOV3 et SW626, l'expression des exons variants représente respectivement environ 5% et 50% de celle des exons constants (**Figure 6A**). Ces deux dernières lignées cellulaires présentent un intérêt particulier puisqu'il a été précédemment montré au laboratoire que l'inclusion des exons variants est corrélée avec les niveaux de méthylation H3K9 et ceux de la protéine de l'hétérochromatine HP1 γ (Saint-André et al., 2011). L'analyse par MeDIP ne révèle pas de différence des patrons de méthylation de l'ADN dans les 3 lignées cellulaires (**Figure 6B**) suggérant qu'à l'intérieur du gène CD44, la méthylation de l'ADN et celle de H3K9 ne coopèrent pas pour réguler l'épissage des exons variants.

4. IMPACT DES CHANGEMENTS DE METHYLATION DE L'ADN SUR L'EPISSAGE ALTERNATIF DE CD44

Les résultats précédents montrent qu'il n'y a pas de corrélations entre les décisions d'épissage de CD44 et les profils de méthylation de l'ADN. Ces résultats amènent des questionnements quant à l'hypothèse de régulation directe de la méthylation de l'ADN sur l'épissage. Néanmoins, l'impact de la méthylation de l'ADN sur la régulation de l'épissage, de manière directe ou indirecte (par exemple, *via* la régulation de l'expression de facteur d'épissage), demeure en suspens. Pour explorer cette hypothèse, nous avons regardé si des traitements visant à réduire la méthylation de l'ADN, affectent l'épissage de CD44.

4.1. Impact du traitement à l'azadeoxycytidine sur l'épissage de CD44

Nous avons traité les cellules HeLa à l'azadeoxycytidine, un inhibiteur de la méthylation de l'ADN. L'efficacité du traitement a été contrôlée en vérifiant l'augmentation de l'expression des ARNm TSH2B, qui code pour un variant d'histone H2B spécifique des testicules dont l'expression est régulée par la méthylation de l'ADN (**Figure 7A**). Le traitement à l'azadeoxycytidine n'affecte pas l'épissage de CD44 (**Figure 7A**). L'analyse par MeDIP nous a permis de mettre en évidence que seules les exons C2 et C5, à l'origine fortement méthylés, présentent une diminution de méthylation de l'ADN dans les cellules traitées à l'azadeoxycytidine (**Figure 7B**). La faible efficacité du traitement à l'azadeoxycytidine est cohérente avec une étude publiée ultérieurement (Hagemann et al., 2011) montrant que cette drogue inhibe la méthylation d'une faible portion des régions méthylées et ceux préférentiellement dans les zones fortement méthylées. Ainsi, cette approche expérimentale est inexploitable dans notre cas.

4.2. Impact de la déplétion des enzymes DNMT sur l'épissage de CD44

Nous avons alors opté pour une stratégie de déplétion par ARN interférence de l'enzyme DNMT1, impliqué dans le maintien de la méthylation de l'ADN au cours de réplication. Les cellules ont été incubées 5 jours post-transfection avant de les récupérer de manière à avoir un taux de déméthylation passif important (déméthylation théorique de 2^5 si on considère que les cellules se divisent toutes les 24 heures).

La déplétion de l'enzyme DNMT1, en transfectant un mélange de 3 petits ARN interférents (siRNA), (**Figure 8A**) dans les cellules HeLa induit une diminution des niveaux de méthylation à l'intérieur du gène CD44, excepté au niveau des exons qui sont peu méthylés (v5, v7, v9 et v10) (**Figure 8B**). Ce résultat est accompagné d'une diminution de l'inclusion des exons de CD44 (**Figure 8C**). Bien que l'approche utilisée ne puisse pas argumenter en faveur d'un rôle direct de la méthylation de l'ADN au locus des exons variants de CD44, nous soulignons que ces résultats sont en accord avec un modèle dans lequel la méthylation de l'ADN affecte négativement l'élongation transcriptionnelle favorisant l'inclusion des exons alternatifs. Néanmoins, il est important de noter que l'un des effets majoritaires se porte sur l'exon v9 alors qu'il ne possède qu'un niveau très faible en méthylation. Tout en restant cohérent avec un mécanisme direct, ces résultats pourraient s'expliquer par le fait que l'inclusion des exons variants de CD44 ne se fait pas indépendamment les uns des autres (Bell et al., 1998). Ainsi, l'inclusion du variant v8, l'exon le plus méthylé avec v4, pourrait favoriser celle de v9.

Par ailleurs, la déplétion de DNMT1 empêche l'inclusion préférentielle des exons variants de CD44 en réponse au traitement PMA (**Figure 8C**).

Ainsi, dans les cellules HeLa, l'enzyme DNMT1 semble responsable de la méthylation à l'intérieur du gène CD44 et semble favoriser l'inclusion des exons variants de CD44 et semble nécessaire à la réponse au PMA.

4.3. Impact de la déplétion des protéines MBD sur l'épissage de CD44

Pour déterminer si l'impact de DNMT1 sur l'épissage de CD44 met en jeu les protéines qui se lient à l'ADN méthylé, nous avons effectué des déplétions des protéines à domaine MBD (**Figure 9A**). Les déplétions de MBD1, MBD2 et MBD3 diminuent l'inclusion de l'ensemble des exons variants de CD44. Celle de MeCP2 engendre une diminution de l'inclusion des exons v2, v4 et v5. Enfin, la déplétion de MBD4 induit une diminution de l'expression de CD44 mais n'a pas d'impact sur son épissage. (**Figure 9B**)

De nouveau, cette approche ne permet pas de distinguer les effets directs des effets indirects mais ensemble, ces résultats suggèrent que DNMT1 et MBD agissent de concert pour réguler l'épissage de CD44.

5. IMPACT DES ENZYMES DNMT SUR L'ÉPISSAGE À L'ÉCHELLE DU GÉNOME

5.1. Impact de la déplétion de l'enzyme DNMT1 dans les cellules HeLa

Par la suite, nous avons décidé d'analyser l'impact de DNMT1 sur l'ensemble du transcriptome. D'une part, une telle approche peut permettre de déterminer si DNMT1 a un effet global sur l'épissage à l'échelle du génome. D'autre part, en vue de comprendre le mécanisme mis en jeu par DNMT1 pour affecter l'épissage, une telle approche peut permettre de déterminer si DNMT1 affecte l'expression de transcrits codant pour des facteurs d'épissage.

Nous avons opté pour l'utilisation de puces exon. Ces puces possèdent des jeux de sondes ciblant toutes les régions potentiellement incluses dans les ARNm. L'analyse des signaux associés à chaque sonde indépendamment les uns des autres permet d'estimer les décisions d'épissage alternatif.

Pour augmenter l'astringence de l'approche, nous avons décidé d'analyser et recouper les informations issues du transcriptome de cellules déplétées de DNMT1 en transfectant individuellement 3 petits ARN interférents (siRNA) différents (siRNA DNMT1_6 ; DNMT1_7 et DNMT1_1). Avec chacun des siRNA, la diminution de la protéine DNMT1 est observée (**Figure 10A**). De plus, leur transfection induit une déméthylation de la région promotrice de H19 (**Figure 10B**), un gène connu pour être régulé par la méthylation de l'ADN, ainsi qu'une dérégulation de ses transcrits (**Figure 10C**). L'analyse de l'épissage de CD44 montre que la déplétion par 2 petits ARN interférent sur 3 induit une diminution de l'inclusion des exons variants de CD44 (**Figure 10D**).

Les analyses des niveaux d'expression globale et de l'épissage, effectuées par GenosplICE (<http://www.genosplICE.com>), ont consisté à comparer (**Figure 10E**):

- la condition siRNA DNMT1_6 à la condition siRNA Ctl (rouge)
- la condition siRNA DNMT1_7 à la condition siRNA Ctl (bleu)
- la condition siRNA DNMT1_1 à la condition siRNA Ctl (vert)

De manière inattendue, le recouvrement des gènes « cibles », pour l'expression globale, de DNMT1, obtenus avec les différents siRNA, est extrêmement faible : entre 2 et 8% en comparant les résultats des siRNA 2 à 2 et 0,7% en prenant en compte les 3 siRNA (**Figure**

10E). En ce qui concerne les événements d'épissage, le recouvrement est quasiment nul (**3 événements communs à 2 siRNA ; non montré**).

A l'issue de ces résultats, de nombreux questionnements sont survenus quant à l'interprétation de cette expérience. La qualité des ARN, celle du marquage et de l'hybridation des puces ont été contrôlé positivement. Par ailleurs, les niveaux d'expression et des événements d'épissage obtenus par RT-qPCR ou par l'analyse des puces sont similaires. Par exemple, la l'augmentation de l'expression de H19 et la diminution de l'expression des exons variants de CD44 avec les siRNA DNMT1_6 et 7 sont observés sur les puces. Nous avons aussi validé des changements d'expression et d'épissage (ou les non changements selon le siRNA utilisé) d'autres gènes « cibles » mis en évidence par l'analyse des puces. Ainsi, nous avons écarté la piste des problèmes expérimentaux et plutôt penché pour un problème biologique.

Evaluation des effets *off-target* potentiels :

Un point a attiré notre attention : à la suite de la déplétion de DNMT1, la plupart des cibles sont réprimées. Pourtant, DNMT1 a principalement été caractérisé pour son implication dans la répression transcriptionnelle. Nous avons donc évoqué la possibilité que les siRNA utilisés engendrent des effets non spécifiques dits «*off-target*», c'est-à-dire que les siRNA pourraient inhiber l'expression d'autres gènes.

Pour tester cette hypothèse, nous avons comparé les effets de la déplétion de DNMT1 en utilisant les précédents siRNA non modifiés chimiquement, ou en utilisant des siRNA de même séquence mais qui sont 2'-o-méthyle à la base 2 du brin antisens dans l'objectif de limiter les effets non spécifiques (Jackson et al., 2006). L'efficacité de ces derniers siRNA est considérablement diminuée, notamment avec le siRNA DNMT1_7 où aucune déplétion de DNMT1 n'est observée (**Figure 11A**). Pourtant, sa transfection induit tout comme son homologue non modifié chimiquement une diminution de l'inclusion des exons variants de CD44 (**Figure 11B**). Ainsi, ce résultat indique que l'impact de la transfection du siRNA DNMT1_7 n'est certainement pas dû à la déplétion de DNMT1.

Plusieurs études ont mis en évidence que la région « *seed* » d'un siRNA, c'est-à-dire les nucléotides 2 à 8 du siRNA, est suffisante pour induire une répression des ARNm (Aleman et al., 2007). Pour aller plus loin dans l'analyse de l'impact des effets non spécifiques (c'est-à-

dire non liés à la déplétion de DNMT1), nous avons effectué une étude bioinformatique pour déterminer si la région *seed* de chaque siRNA est présente dans les ARNm des gènes réprimés après transfection.

Concrètement, nous avons évalué le pourcentage des gènes réprimés par la transfection du siRNA qui contiennent la région *seed* du siRNA utilisé dans la séquence de l'ARNm. En tant que contrôle, nous avons évalué le pourcentage de ces mêmes gènes qui contiennent la région *seed* des 2 autres siRNA non utilisés. Dans chacun des cas, le pourcentage obtenu (% observé) a été comparé au pourcentage théorique (% théorique) des gènes contenant la région *seed* (Nombre de transcrits décrits dans hg19 contenant la région *seed*/ Nb de transcrits total décrits dans hg19 *100). De manière intéressante, le pourcentage observé de gènes cibles contenant la région *seed* du siRNA utilisé est supérieur au pourcentage théorique et ce, de manière plus importante qu'avec la région *seed* des siRNA non utilisés. **(Figure 11C)**

Ainsi, l'effet *target off* des siRNA pourraient expliquer en partie l'absence de recouvrement entre nos différents résultats.

Amplification des effets liés aux transfections sur 5 jours :

Par ailleurs, les transfections ont été effectuées sur 5 jours (ayant pour objectif de réduire significativement les niveaux de méthylation de l'ADN). Ainsi, les éventuels effets non spécifiques induit par la transfection des siRNA ont pût être amplifiés pendant les 5 jours d'incubation post-transfection.

Ainsi, la distinction des effets spécifiques et non spécifiques de la déplétion de DNMT1 étant impossible, les résultats issus des transfections des siRNA DNMT1 sont difficilement interprétables.

5.2. Impact de l'invalidation des enzymes DNMT1, DNMT3A et DNMT3B dans les cellules ES

5.2.1. Analyse du transcriptome des cellules ES invalidées pour les DNMT

Pour s'affranchir de l'utilisation de siRNA, nous avons entrepris d'analyser des données de séquençage à haut débit du transcriptome de cellules souches embryonnaires murines

invalidées pour l'expression de DNMT1, 3A, 3B (TKO) et des cellules parentales J1 (WT) (Karimi et al., 2011).

L'analyse montre que la proportion des gènes induits *versus* réprimés est similaire : 161 gènes sont surexprimés dans les cellules TKO et 142 sont réprimés (**Figure 12A**). Par ailleurs, le facteur de changement des gènes augmentés est plus important que celui des gènes réprimés suggérant qu'une plus grande proportion de gènes surexprimés provient d'un effet direct de la déplétion des enzymes DNMT imposant une répression transcriptionnelle (**Figure 12A**).

Concernant l'épissage, l'analyse du RNAseq a mis en évidence 289 exons cassettes différentiellement inclus : 112 sont plus inclus et 177 sont moins inclus dans les cellules TKO comparées aux cellules WT (**Figure 12B**).

De manière intéressante, les groupes de gènes différentiellement exprimés et ceux différentiellement épissés ne sont pas les mêmes (**Figure 12C**). Ainsi, les résultats suggèrent que l'impact des enzymes DNMT sur l'épissage ne provient pas des promoteurs et des facteurs de transcription recrutés.

Nous avons aussi regardé si l'inactivation des DNMT affecte l'expression de facteurs d'épissage. L'analyse a révélé le changement d'expression d'un seul facteur : Mbnl1 qui est surexprimé près de 6 fois dans les cellules TKO. Mbnl1 affecte l'épissage de 103 gènes dans des cellules ES murines (Han et al., 2013). Seul 10 de ces gènes sont retrouvés différemment épissés dans les cellules TKO. Ainsi, cette comparaison suggère que l'effet de l'inactivation des DNMT sur l'épissage ne provient pas uniquement de la régulation de l'expression de Mbnl1.

5.2.2. Analyse du méthylome des cellules ES

L'équipe de W. Reik, a analysé le méthylome par MeDIP suivi de séquençage à haut débit des cellules ES J1 (Ficz et al., 2011). L'analyse de leurs données, par Genosplice, nous a permis de déterminer l'intensité de pics de méthylation dans les exons cassettes régulés (ceux déterminés précédemment), dans les exons non régulés des mêmes gènes et dans les exons non régulés de tous les gènes. Aucune différence significative n'a été observée (**Figure 13A**). En revanche, les exons régulés possèdent plus de pics de méthylation que les exons non régulés des mêmes gènes (**Figure 13B**) et aussi par rapport à l'ensemble des exons non régulés (**Figure 13C**).

Ainsi, ces résultats apportent le premier argument, corrélatif, en faveur d'un modèle où la méthylation de l'ADN pourrait affecter l'épissage de manière directe, c'est-à-dire au locus de l'exon régulé.

6. CONCLUSIONS

Le but de cette étude est de déterminer si la méthylation de l'ADN a un rôle dans la régulation de l'épissage alternatif et d'en définir le mécanisme.

Bilan sur les résultats obtenus avec le gène modèle CD44 :

Nos résultats issus de la déplétion de DNMT1 dans les cellules HeLa sont inexploitable. En revanche, des résultats obtenus par Benjamin Hopkins au laboratoire montrent que les enzymes DNMT régulent l'épissage de CD44. En effet, dans les cellules HCT116 doublement invalidées pour DNMT1 et DNMT3B (Rhee et al., 2002), les exons variants de CD44 sont moins inclus que dans les cellules parentales. De plus, la déplétion des protéines à domaine MBD1 MBD2 et MBD3 induisent aussi une diminution de l'inclusion des exons variants de CD44. Ainsi, les enzymes DNMT et les protéines MBD pourraient agir de concert pour réguler l'épissage de CD44.

Par ailleurs, nous avons mis en évidence avec différents systèmes d'études que les décisions d'épissage de CD44 et les niveaux de méthylation de l'ADN intragéniques ne sont pas corrélés. Ensemble, ces résultats suggèrent plutôt que les enzymes DNMT régulent l'épissage de CD44 par un mécanisme indirect.

Bilan sur les résultats obtenus à l'échelle du génome :

L'analyse du transcriptome des cellules ES murine invalidées pour les enzymes DNMT a permis de mettre en évidence un impact global de ces enzymes sur l'épissage. De manière intéressante, les exons régulés sont enrichis en méthylation de l'ADN dans les cellules parentales suggérant que la méthylation de l'ADN pourrait affecter l'épissage de manière directe.

Par ailleurs, nous avons mis en évidence que les exons définis comme cassettes d'après la base de données *Genome Browser* sont appauvris en dinucléotides CpG. Il sera intéressant de

déterminer si les exons cassettes régulés dans les TKO sont aussi appauvris en CpG. Si tel est le cas, les résultats pourraient être la base d'un modèle où des exons cassettes compensent leur faible taux en dinucléotides CpG en augmentant la proportion de CpG méthylés.

Néanmoins, cette même étude met en évidence que l'expression du facteur d'épissage Mbnl1 est régulé par l'invalidation des DNMT suggérant aussi que la régulation de l'épissage puisse être une conséquence de la régulation de l'expression de facteur d'épissage.

Conclusion :

La somme de tous ces résultats illustre la complexité de l'interprétation des différentes approches utilisées. Toutes sont basées sur les comparaisons génomiques et transcriptomiques de systèmes présentant de nombreuses variables. Dans cette étude, nous avons montré que les enzymes DNMT affectent l'épissage de nombreux gènes mais la compréhension du mécanisme mis en jeu nécessitera le développement de nouveaux outils.

Bilan sur les publications parues lors de ma thèse :

Deux publications clefs traitant des éventuels liens entre la méthylation de l'ADN et les décisions d'épissage.

- En 2011, S. Shukla *et al.* ont mis évidence que la protéine CTCF, une protéine qui se lie aux séquences ADN CCCTC préalablement décrite pour son rôle dans le ciblage des insulateurs, favorise l'inclusion de l'exon alternatif 5 de CD45 (Shukla et al., 2011). En se liant à un site dans l'exon 5 (de l'ADN), CTCF crée un « barrage » et induit une pause de l'ARN polymérase II. L'impact de CTCF sur la vitesse de la polymérase II a été montré par une approche de transcription *in vitro* d'une matrice d'ADN contenant un site de liaison à CTCF. Ce type d'approche établit un argument fort quand au potentiel effet de CTCF sur l'épissage *via* un modèle cinétique. De plus, la liaison de CTCF à l'ADN est inhibée lorsque que l'ADN est méthylé (Renda et al., 2007). En établissant des corrélations entre les niveaux de méthylation de l'ADN, la présence de CTCF dans l'exon 5 et le patron d'inclusion de cet exon dans l'ARNm de CD45, les auteurs ont suggéré que la méthylation de l'ADN régule négativement l'inclusion de l'exon 5 de CD45 en empêchant la liaison de CTCF au locus génique de l'exon 5.

Néanmoins, dans notre étude l'invalidation de DNMT1 et DNMT3B induit une diminution de l'inclusion des exons variants de CD44 ce qui est en sens opposé de la

publication de S. Shukla *et al.* ; bien que comme évoquées dans l'introduction, les conséquences sur l'épissage d'un ralentissement ou une pause de l'ARN II dépendent du type de facteurs d'épissage recrutés. Par ailleurs, d'après les données de la base de données *Genome Browser*, le facteur CTCF n'est pas présent dans le groupe d'exons variants du gène CD44 dans les cellules HeLa. Ainsi, il semble peu probable que l'impact de l'inactivation de DNMT1 et DNMT3B sur l'épissage soit directement lié à CTCF.

- En 2013, A. Maunakea *et al.*, ont publié une étude dans laquelle ils suggèrent que la méthylation de l'ADN à l'intérieur des gènes module l'épissage *via* le recrutement de MeCP2 (Maunakea et al., 2013). Cependant, cette étude présente de nombreuses faiblesses liées aux approches expérimentales utilisées pour appuyer leur modèle.

L'étude est basée sur deux systèmes d'études :

- le traitement ou non à l'azadeoxycytidine (AZAdC) dans des cellules IMR90,
- le système cellulaire avec les cellules HCT116 doublement invalidées pour DNMT1 (DKO) et DNMT3B (Rhee et al., 2002) et les cellules sauvages (WT).

Corrélations entre méthylation de l'ADN et épissage alternatif ?

En utilisant des données transcriptomiques à grande échelle (RNA-seq), les auteurs montrent que plusieurs centaines d'événements sont régulés dans les 2 systèmes mais ils n'ont pas évalué le recouvrement des événements régulés selon les 2 approches.

Les auteurs se focalisent uniquement sur une sous population d'exons : ceux qui sont inclus à plus de 90% ou moins à de 10% dans les cellules IMR90 non traitées à l'AZAdC, ou dans les cellules HCT116 WT. De plus, les auteurs se basent seulement sur leurs données de séquençage pour évaluer l'épissage (jamais de RT-PCR pour valider les événements étudiés) qui parfois reposent sur seulement 3 *reads* de séquençage! Ainsi, il est probable que certains événements d'épissage alternatif détectés soient le fruit de biais expérimentaux.

A. Maunakea *et al.* montrent sur un seul gène cible, HAUS8, la diminution de la méthylation dans l'exon 10 (l'exon alternatif) dans les cellules HCT116 DKO *versus* WT. Les auteurs n'ont jamais évalué les niveaux de méthylation intragénique à la suite du traitement à l'AZAdC alors qu'ils citent une référence pointant l'inefficacité de cette drogue (Mossman et al., 2010).

Implication de MeCP2 dans la régulation de l'épissage alternatif ?

Les auteurs montrent par ChIP que MeCP2 est présent sur l'exon 10 de HAUS8 sur les cellules HCTT16 WT et les IMR90 non traitées à l'AZAdC mais n'est pas présent dans les cellules HCT116 DKO et les IMR90 traitées à l'AZAdC. En revanche, A. Maunakea *et al.* ne montrent pas l'impact de la déplétion de MeCP2 sur l'épissage de HAUS8 alors qu'ils ont des données de RNA-seq dans les cellules IMR90 et HCT116 WT dont l'expression de MeCP2 est inhibée par ARN interférence.

Pour conclure, cette étude n'est que peu informative à propos du rôle potentiel de la méthylation de l'ADN à l'intérieur des gènes sur la régulation de l'épissage. La poursuite de notre étude, à savoir si la méthylation intragénique peut affecter l'épissage de manière directe par un mécanisme non lié à CTCF, reste alors une question ouverte pour la communauté du couplage épissage - chromatine. La compréhension du mécanisme mis en jeu nécessitera le développement de nouveaux outils.

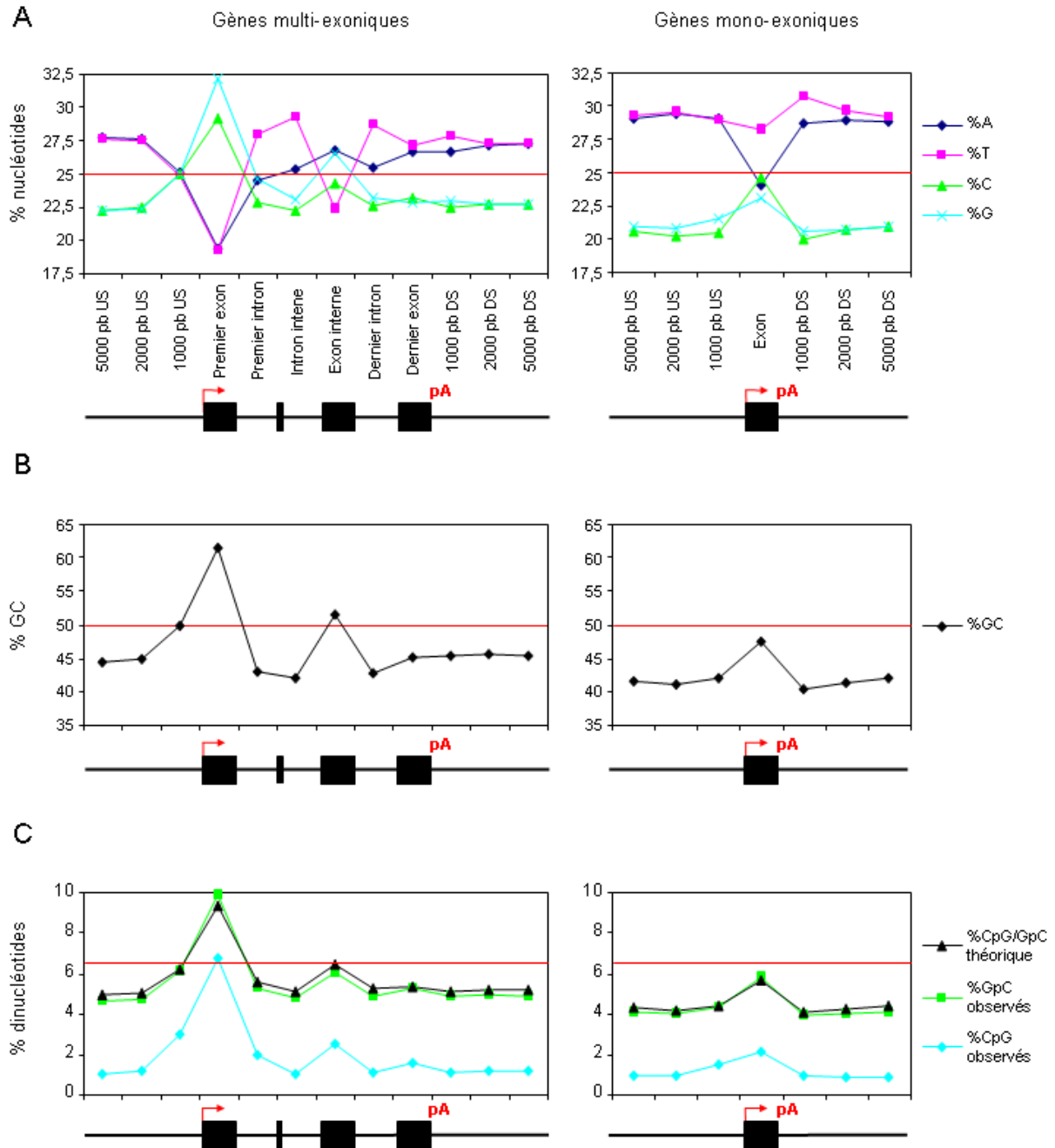


Figure 1. Biais des proportions nucléotidiques dans les différents éléments d'un méta-gène de souris. A. Pourcentage de chaque nucléotide dans les différents éléments d'un méta-gène multi-exonique (panneau gauche) ou mono-exonique (panneau droite). **B.** Pourcentage de GC dans les différents éléments d'un méta-gène multi-exonique (panneau gauche) ou mono-exonique (panneau droite). **C.** Pourcentage observés et théoriques (en tenant compte des proportions en G et en C dans chaque éléments du méta-gène) des dinucléotides CpG et GpC dans les différents éléments d'un méta-gène multi-exonique (panneau gauche) ou mono-exonique (panneau droite). Dans chaque panneau la ligne rouge représente les pourcentages de (di-)nucléotides théoriques si les 4 nucléotides étaient en proportions équivalentes et répartis de manière homogène dans le génome. 1000 pb US/DS = région de 1 pb à 1000 pb en amont ou aval respectivement ; 2000 pb US/DS = région de 1001 pb à 2000 pb ; 5000 pb US/DS = région de 2001 pb à 5000 pb.

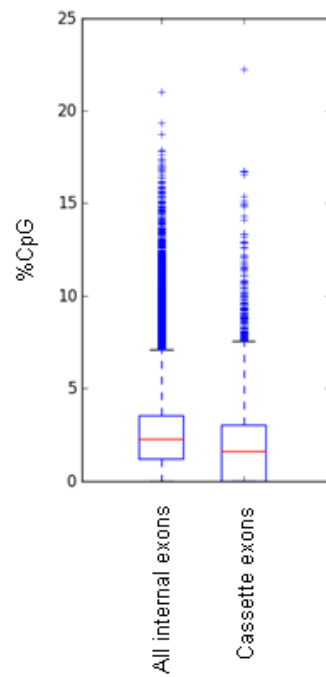


Figure 2. Les cassettes exons sont moins riches en dinucléotides CpG. Le pourcentage de dinucléotides CpG a été évalué dans l'ensemble des exons internes du génome de souris et dans les exons cassettes définis comme cassettes dans Genome Browser.

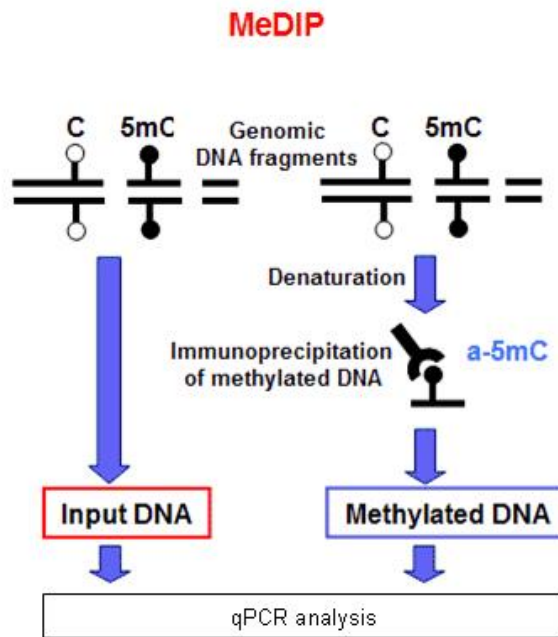
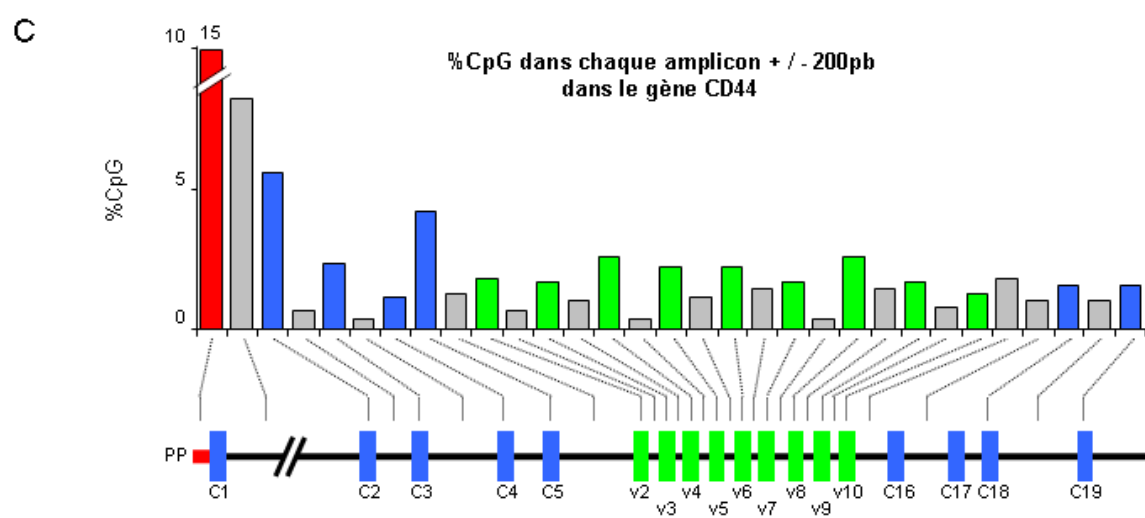
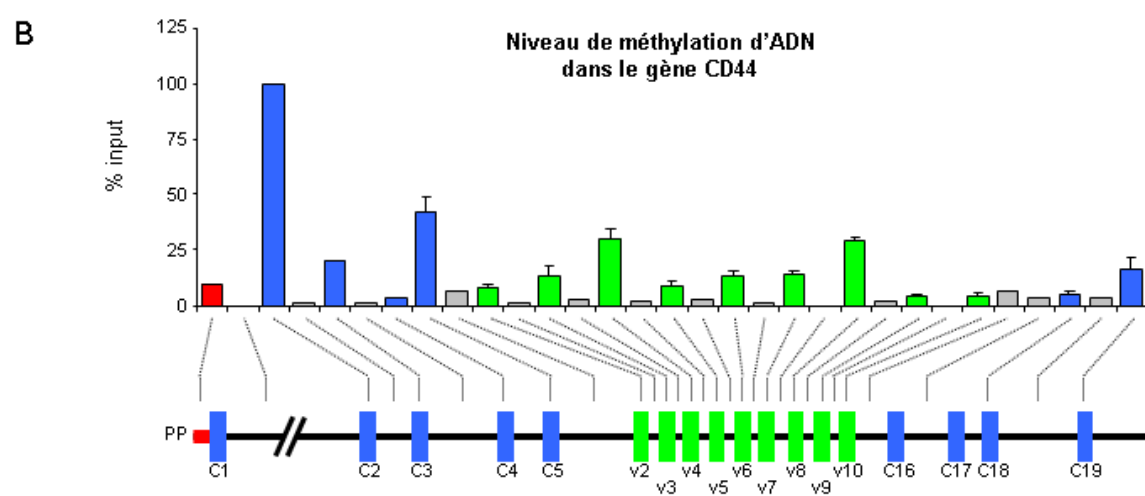
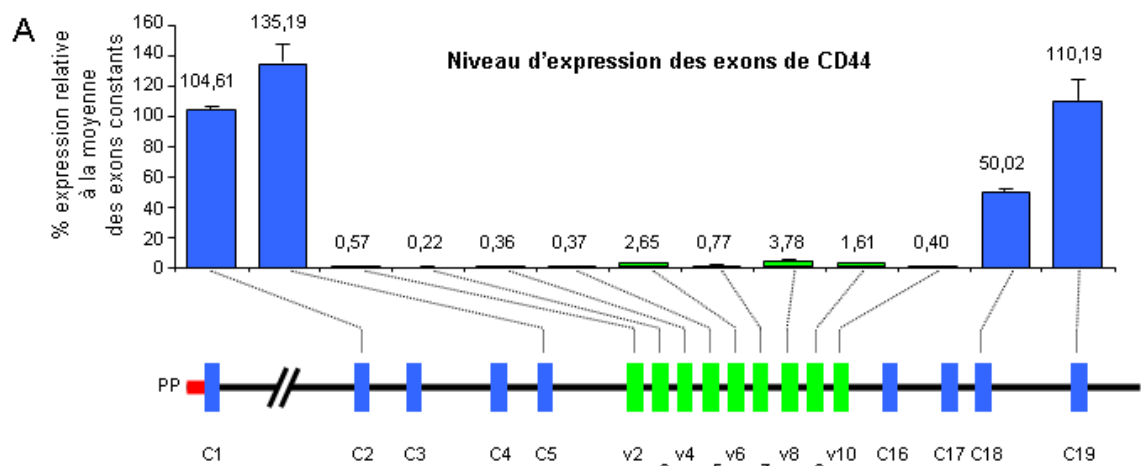


Figure 3. Schéma explicatif de la procédure du MeDIP. Après avoir extrait et soniqué l'ADN des cellules, cette expérience consiste à immunoprécipiter l'ADN méthylé en utilisant un anticorps anti 5-méthyl-cytosine (5-mC). L'enrichissement relatif de méthylation de l'ADN est analysé par PCR quantitative (qPCR).



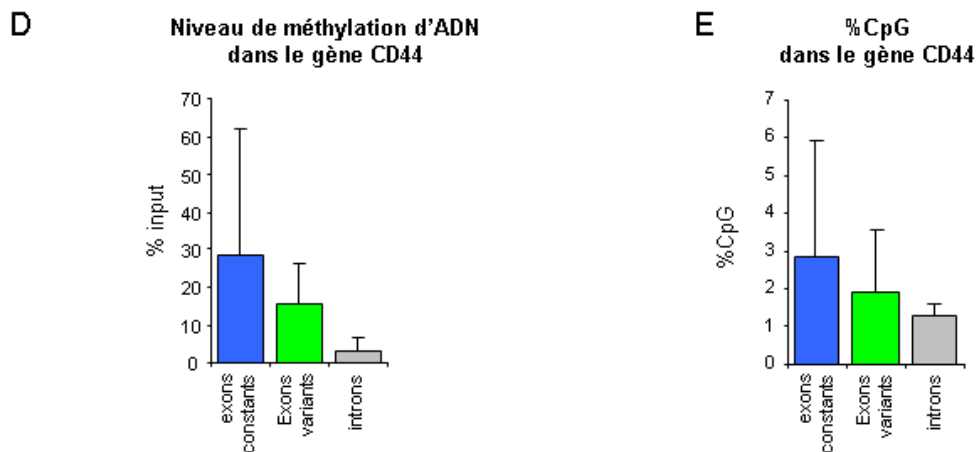


Figure 4. Les exons du gène CD44 sont enrichis en méthylation de l'ADN. **A.** Expression des exons de CD44 relative à l'expression moyenne des exons constants (C2, C5, C18 et C19) dans les cellules HeLa. La région promotrice proximale « PP » est associée au rouge, les exons constants « C » au bleu, les exons variants « v » au vert et les introns au gris. **B.** Analyse par MeDIP-qPCR du niveau de méthylation relatif à l'input dans les exons et les introns du gène CD44 dans les cellules HeLa. **C.** %CpG dans les séquences [A-200pb;A+200pb] où A=amplicons amplifiés par PCR. **D.** et **E.** Moyennes et écart types des niveaux de méthylation (**D**) et du %CpG (**E**) dans les exons constants, variants et les introns du gène CD44.

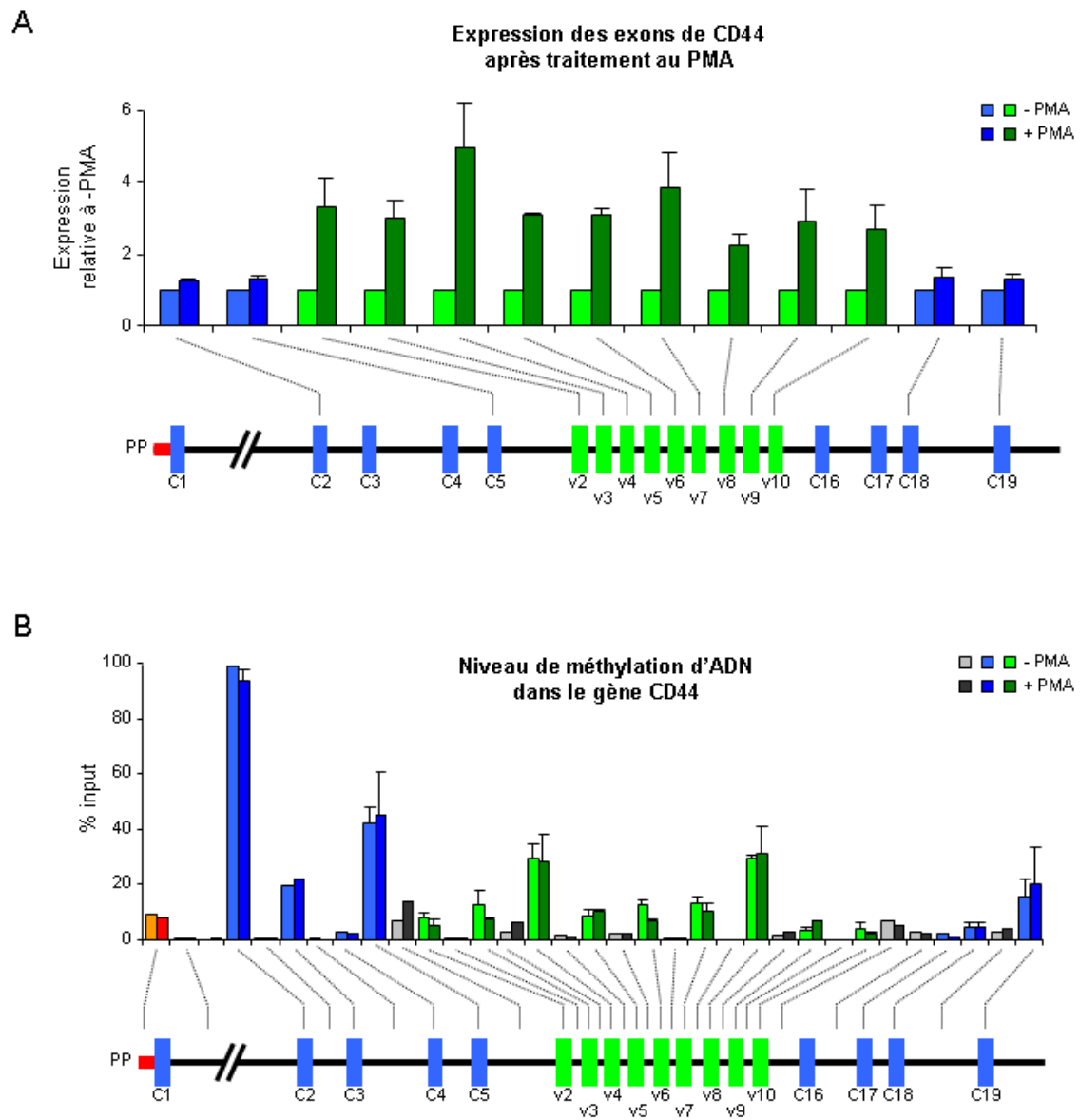
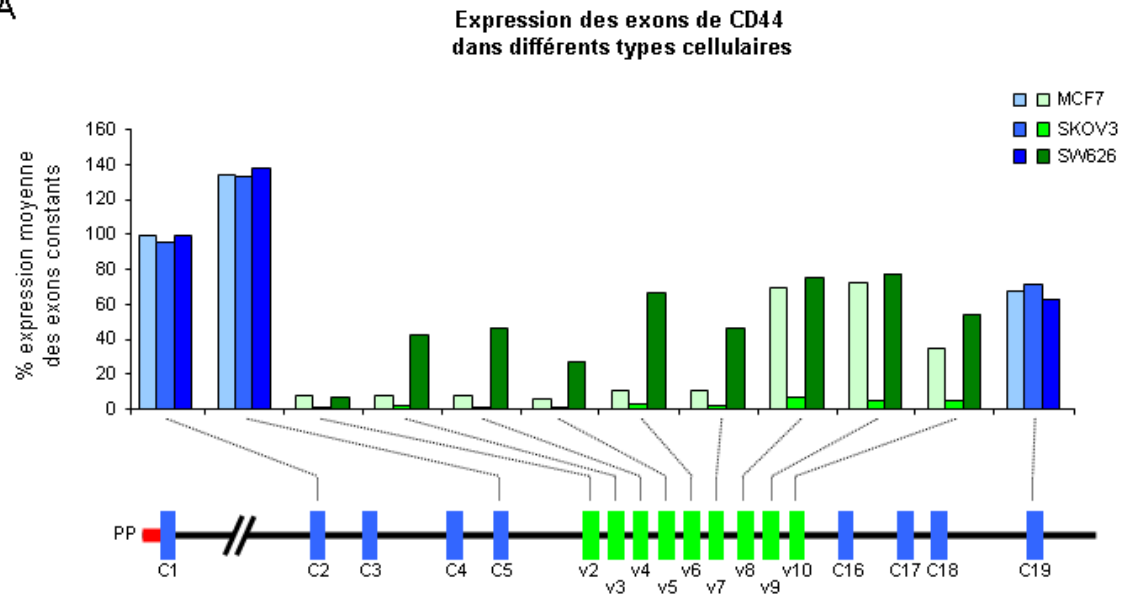


Figure 5. Le traitement au PMA n'affecte pas la méthylation de l'ADN à l'intérieur du gène CD44. A. Expression des exons de CD44 dans les cellules HeLa après 4 heures de traitement au PMA (+PMA) relative à leur expression sans traitement (-PMA). **B.** Niveau de méthylation de l'ADN à l'intérieur du gène CD44 dans les cellules HeLa traitées (+PMA) ou non (-PMA) pendant 4 heures.

A



B

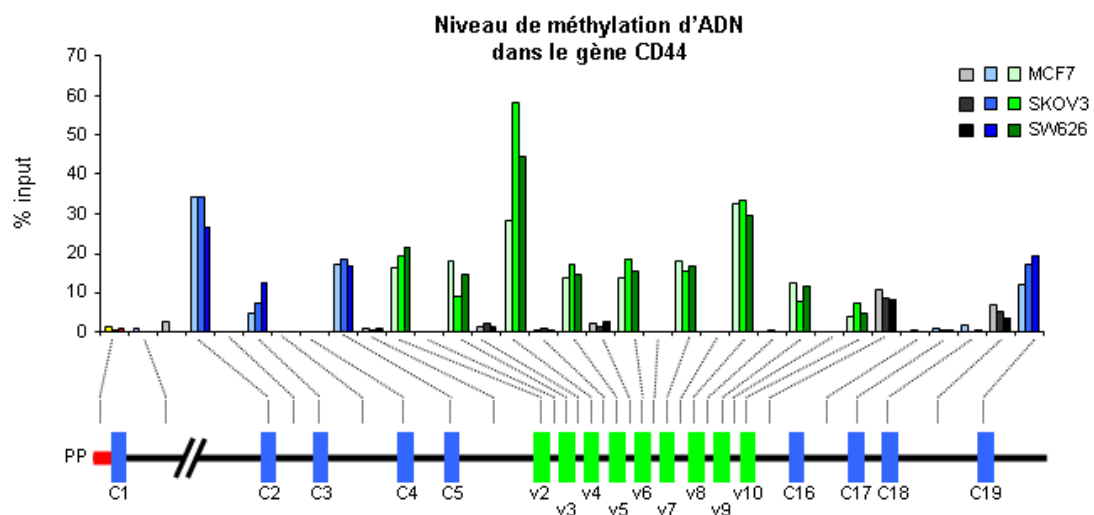


Figure 6. Les profils de méthylation de l'ADN à l'intérieur du gène CD44 ne sont pas corrélés aux profils d'épissage de CD44 dans les cellules MCF7, SKOV3 et SW626. A. Expression des exons de CD44 relative à l'expression moyenne des exons constants (C2, C5, C19) dans les cellules MCF7, SKOV3 et SW626. **B.** Niveau de méthylation de l'ADN à l'intérieur du gène CD44 dans les cellules MCF7, SKOV3 et SW626.

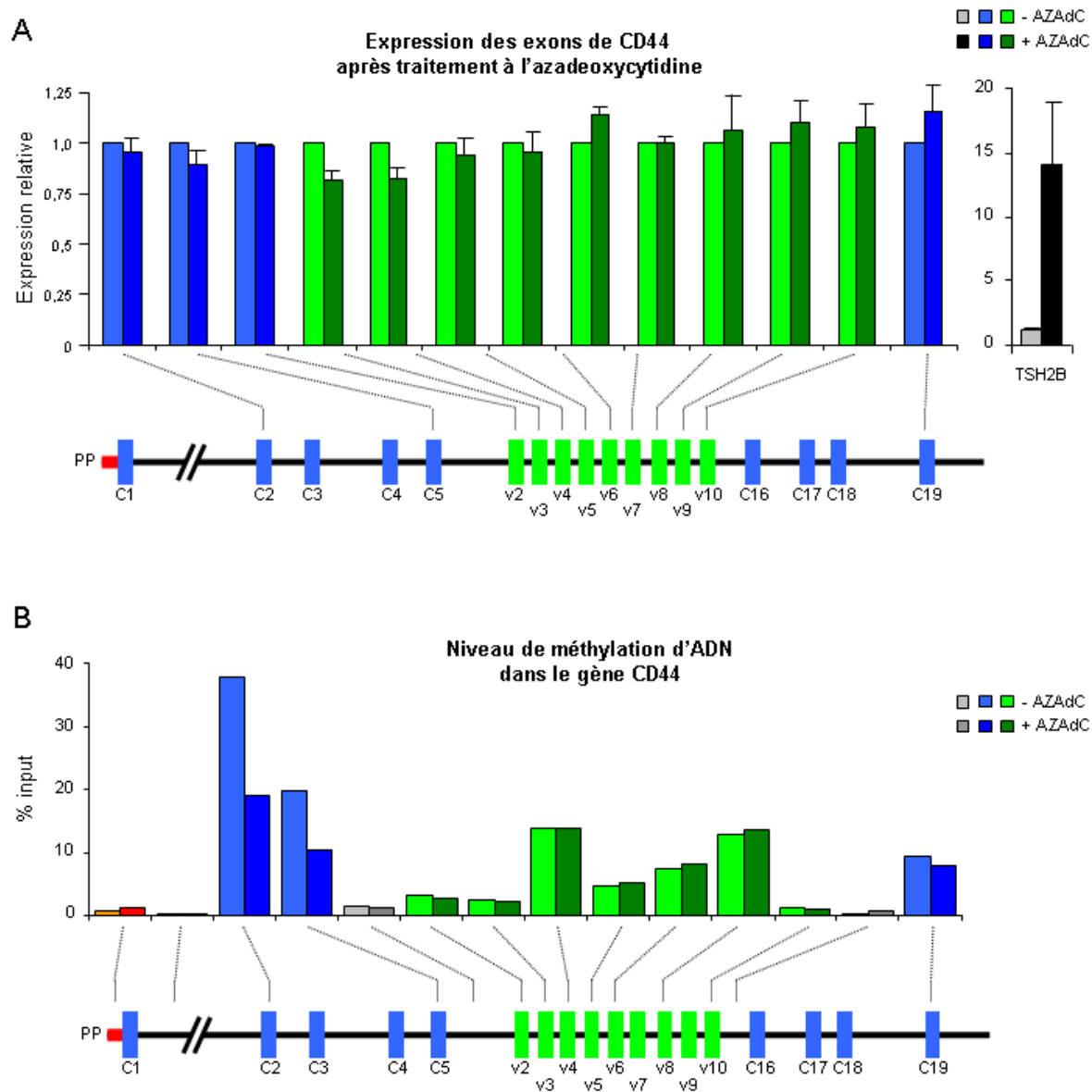


Figure 7. Le traitement à l'azadeoxycytidine est inefficace sur la déméthylation intragénique de CD44. A. Expression des exons de CD44 après traitement à l'azadeoxycytidine (+AZAdC) dans les cellules HeLa relative à leur expression dans des cellules non traitées (-AZAdC). L'expression de TSH2B est affichée comme un contrôle de l'efficacité du traitement. **B.** Niveau de méthylation de l'ADN à l'intérieur du gène CD44 dans les cellules HeLa traitées (+AZAdC) ou non à l'azadeoxycytidine (-AZAdC).

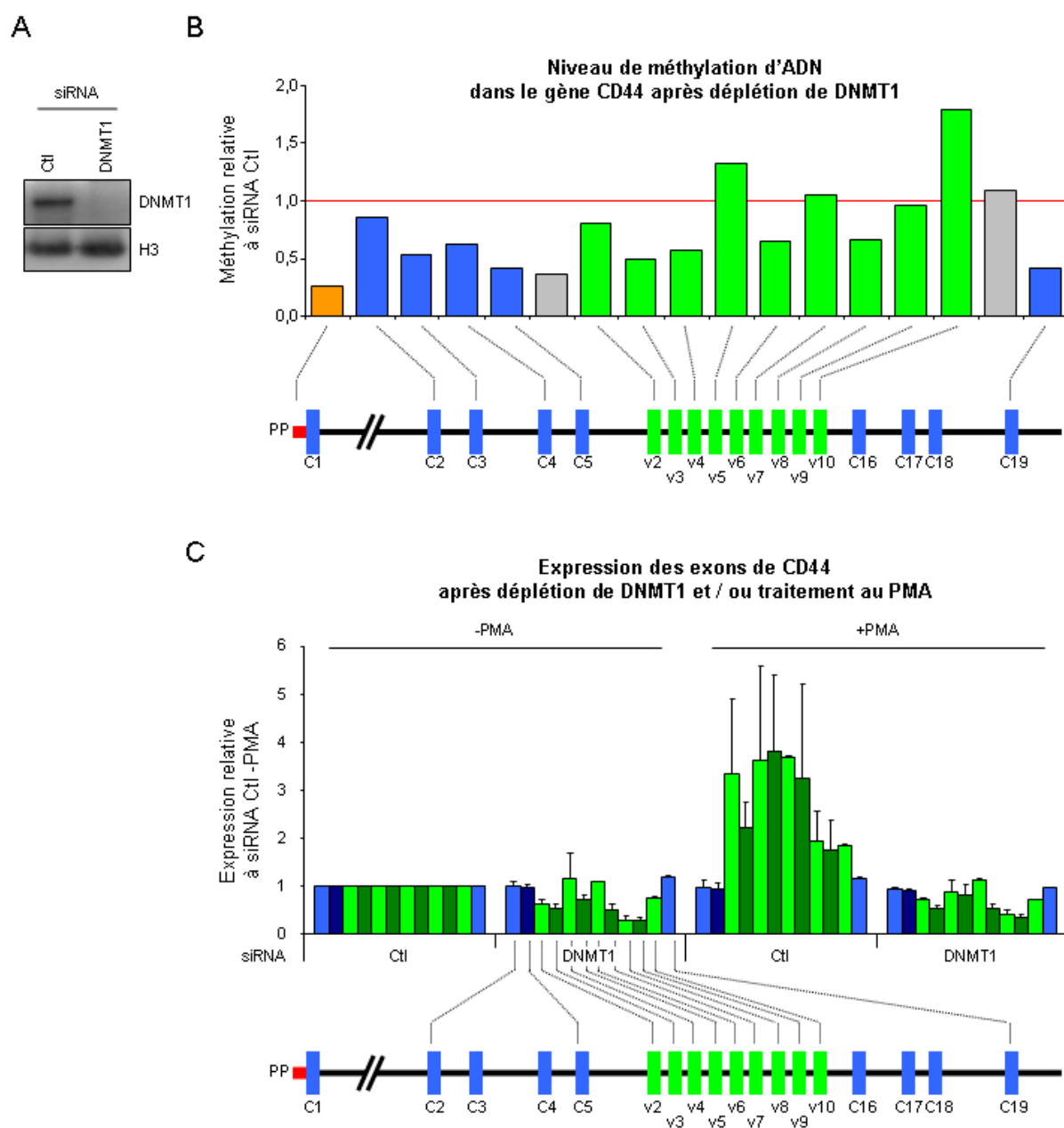


Figure 8. La déplétion de DNMT1 induit une diminution de l'inclusion des exons variants de CD44. A. Des extraits protéiques totaux, issus des cellules HeLa transfectées par des siRNA dirigés contre DNMT1 ou contrôle (Ctl), ont été immunomarkés avec un anticorps dirigé contre DNMT1. Le marquage des histones H3 est utilisé comme contrôle de charge. B. Niveau de méthylation de l'ADN à l'intérieur du gène CD44 dans les cellules HeLa après déplétion de DNMT1 relatif au contrôle. C. Expression des exons de CD44 après déplétion de DNMT1 et/ou traitement au PMA relative à l'expression obtenue dans des cellules HeLa transfectées par le siRNA contrôle (Ctl) non traitées au PMA.

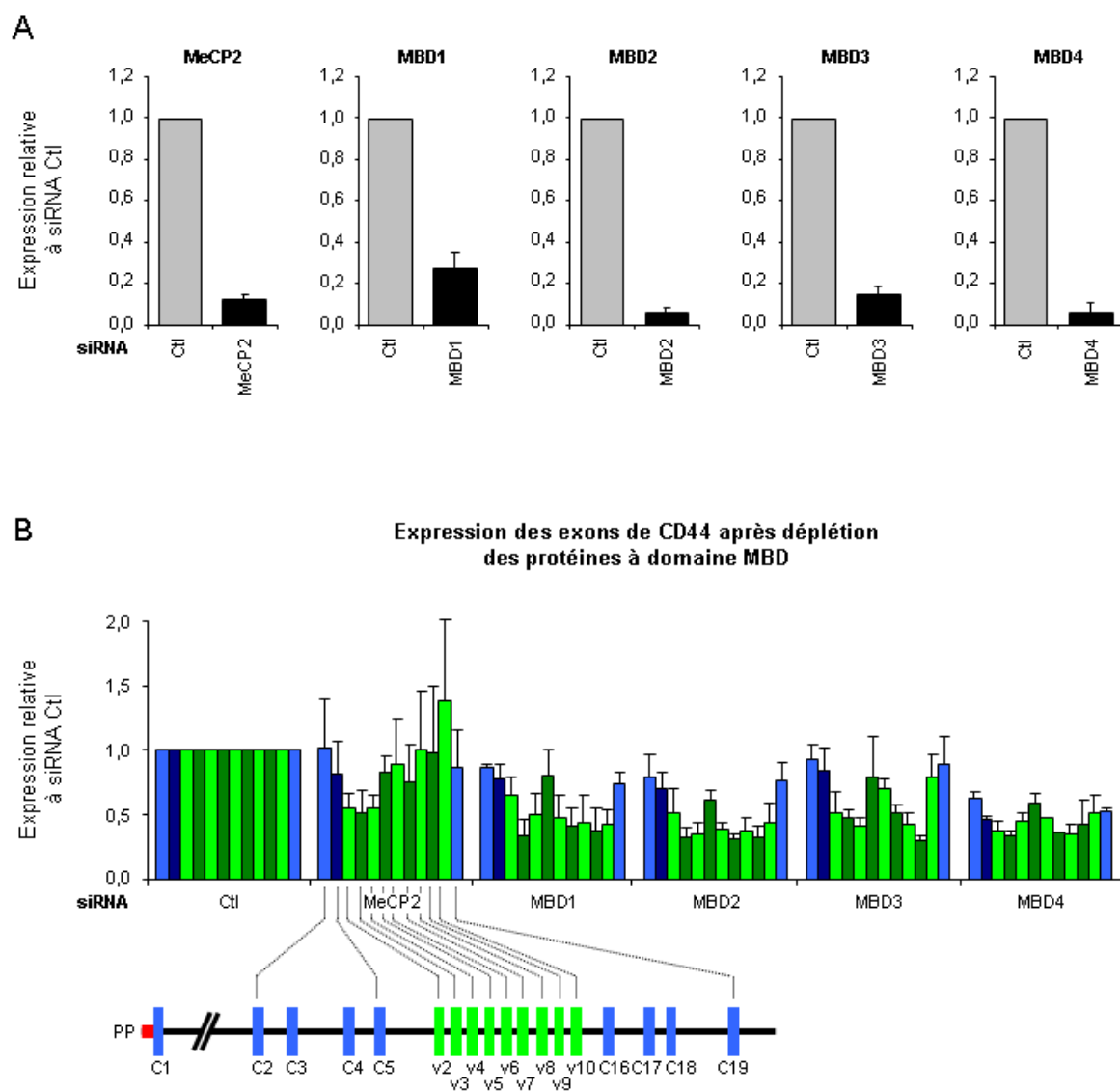


Figure 9. La déplétion des protéines à domaine MBD affecte l'épissage des exons variants de CD44. A. Expression des transcrits codant pour les protéines à domaine MBD relative à l'expression obtenue dans des cellules HeLa transfectées par le siRNA contrôle (Ctl). **B.** Expression des exons de CD44 après déplétion des protéines à domaine MBD relative à l'expression obtenue dans des cellules HeLa transfectées par le siRNA contrôle (Ctl).

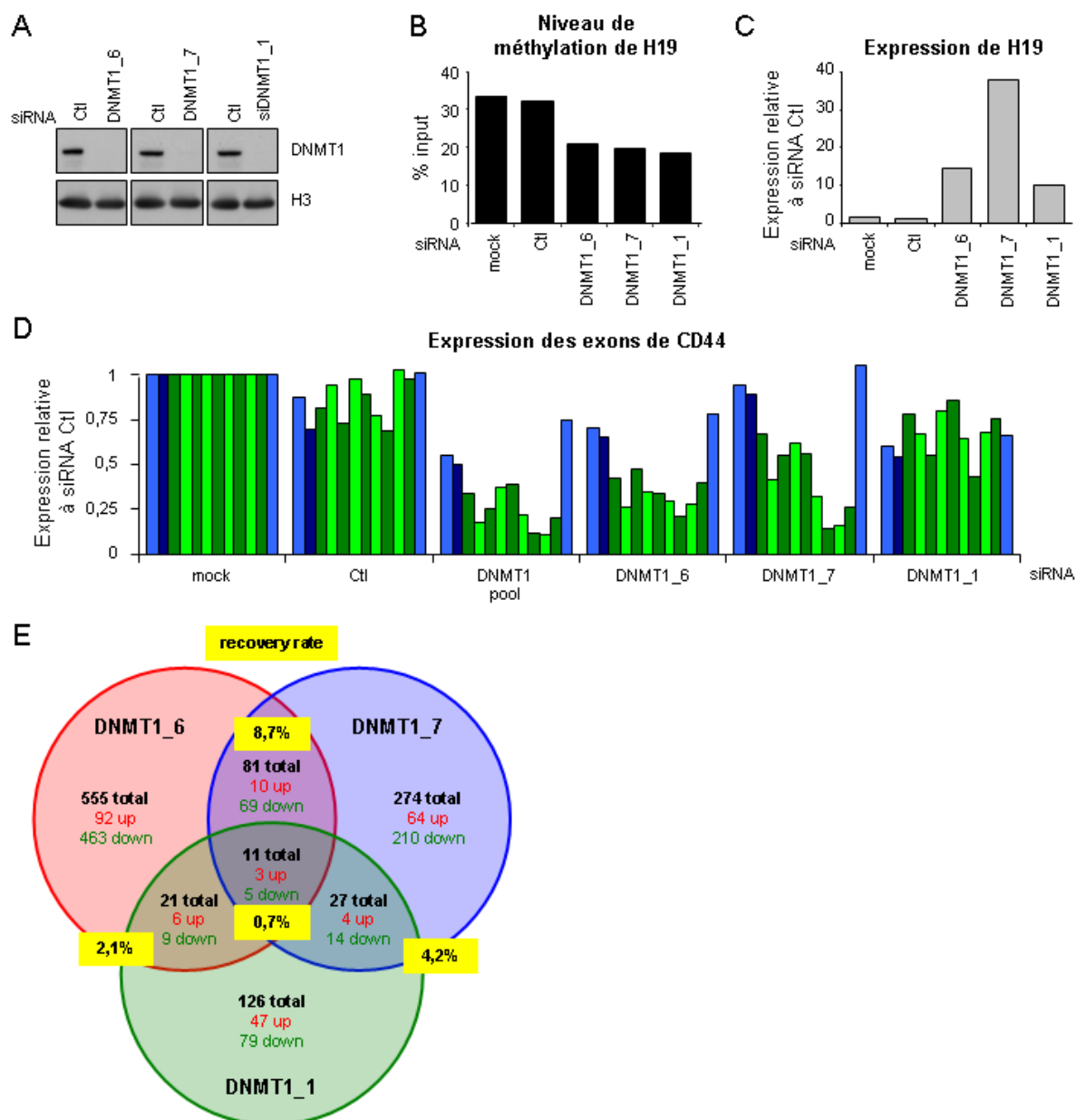


Figure 10. Les siRNA ciblant DNMT1 ont des effets différents. **A.** Des extraits protéiques totaux, issus des cellules HeLa transfectées par des siRNA dirigés contre DNMT1 ou contrôle (Ctl), ont été immunomarqués avec un anticorps dirigé contre DNMT1. Le marquage des histones H3 est utilisé comme contrôle de charge. **B.** Niveau de méthylation de l'ADN dans la région promotrice de H19 dans les cellules HeLa après déplétion de DNMT1 relatif au contrôle. **C.** Expression de H19 après les déplétions de DNMT1 dans les cellules HeLa relative à l'expression dans les cellules contrôle (Ctl). **D.** Expression des exons de CD44 après déplétion de DNMT1 relative à l'expression obtenue dans des cellules HeLa transfectées par le siRNA contrôle (Ctl). **E.** Bilan des gènes différentiellement régulés par les déplétions de DNMT1 en fonction du siRNA utilisé. Les pourcentages de recouvrement sont affichés dans les encarts jaunes.

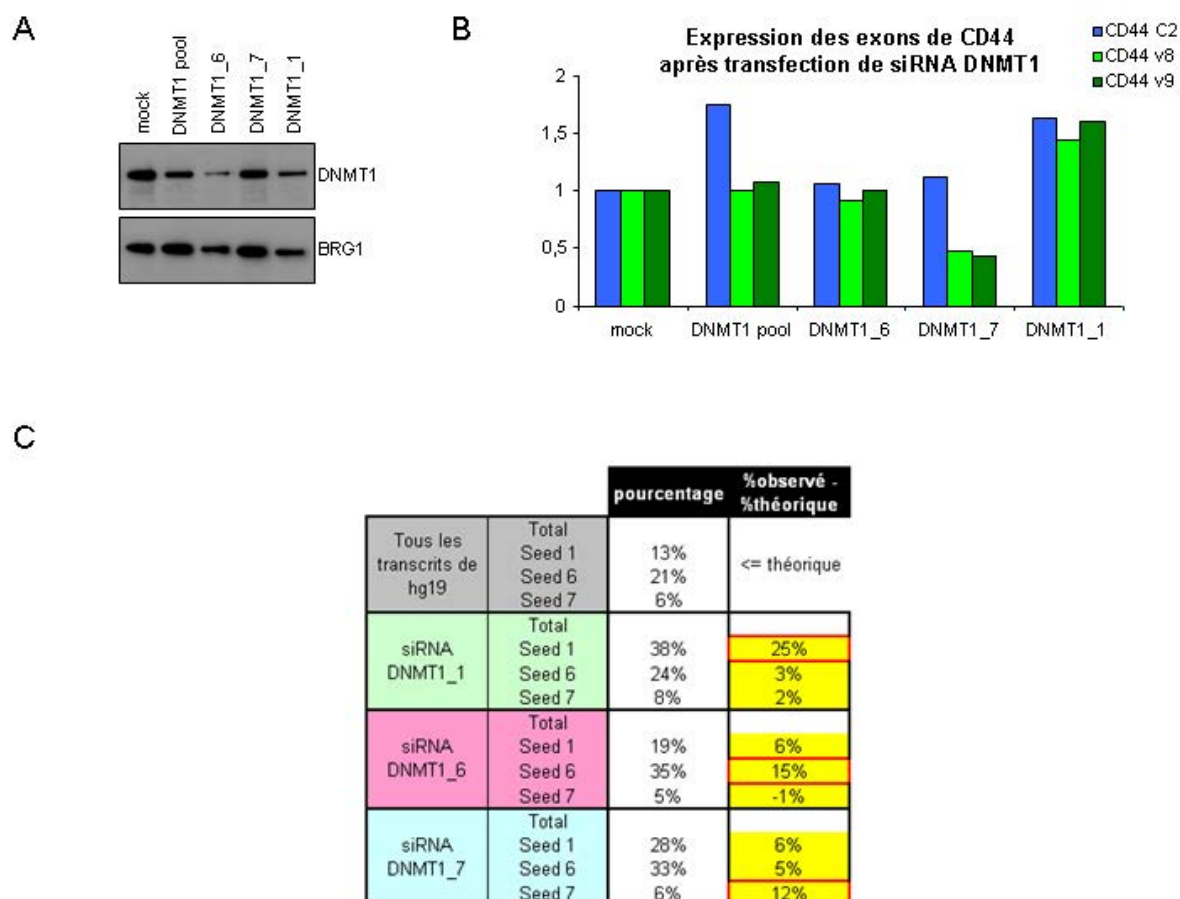


Figure 11. Les siRNA ciblant DNMT1 ont des effets non spécifiques. **A.** Des extraits protéiques totaux, issus des cellules HeLa transfectées par des siRNA 5'-o-methyl dirigés contre DNMT1 ou contrôle (Ctl), ont été immunomarkés avec un anticorps dirigé contre DNMT1. Le marquage Brg1 est utilisé comme contrôle de charge. **B.** Expression des exons de CD44 après la transfection de siRNA 5'-o-methyl dirigés contre DNMT1 dans les cellules HeLa relative à l'expression dans les cellules contrôle (Ctl). **C.** Analyse du pourcentage des cibles réprimés après la transfection des siRNA DNMT1 contenant la région *seed* du siRNA et comparaison (%observé - %théorique) avec les pourcentages de transcrits du génome humain (hg19) contenant les régions seed des différents siRNA.

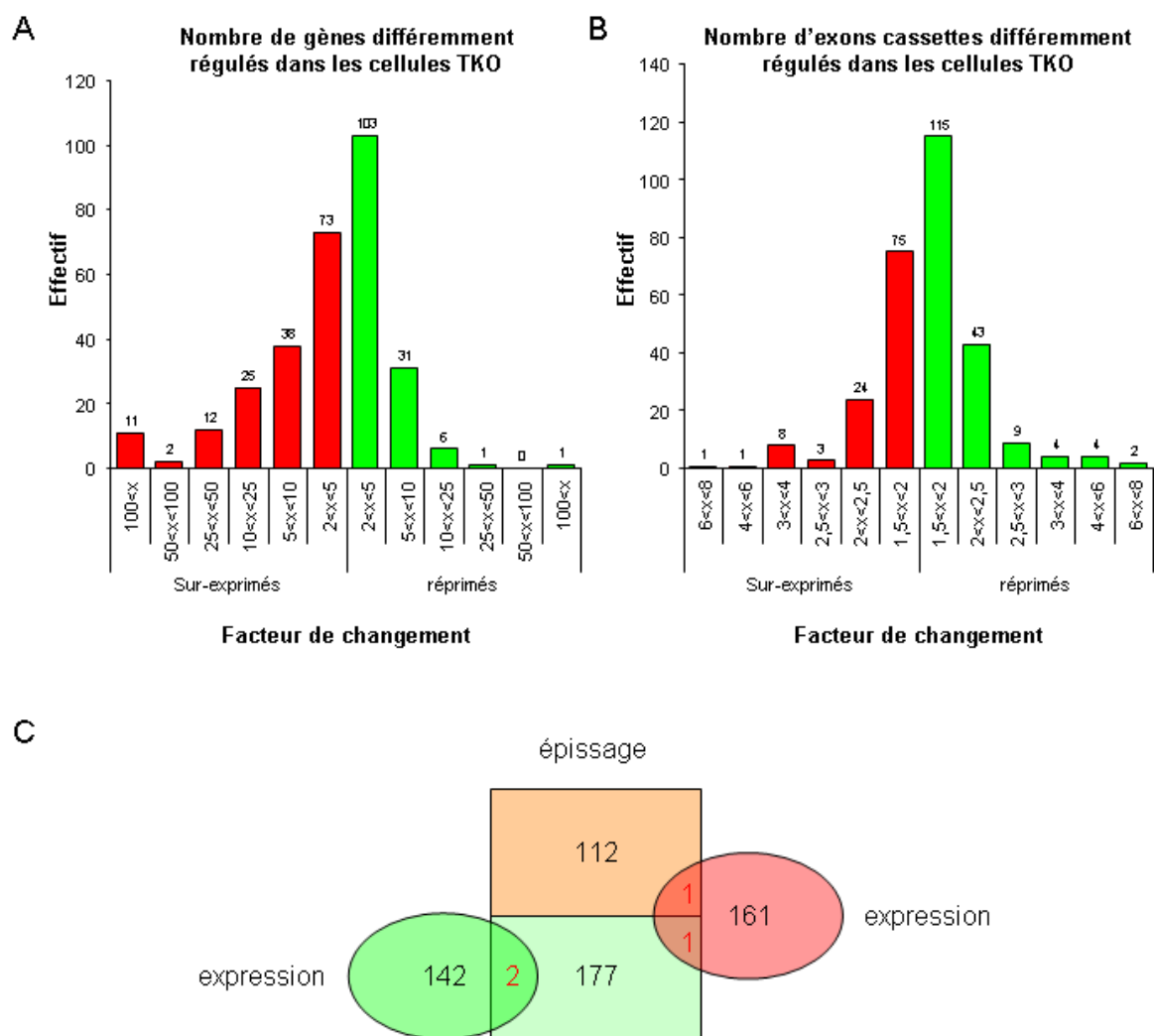


Figure 12. L'inactivation des enzymes DNMT1, DNMT3A et DNMT3B affecte le transcriptome des cellules ES murines. A. Nombres de gènes différemment exprimés dans les cellules ES invalidées pour les DNMT (TKO) comparées aux cellules parentales (WT) en fonction du facteur de changement d'expression. **B.** Nombres d'exons cassettes différemment régulés dans les cellules ES invalidées pour les DNMT (TKO) comparées aux cellules parentales (WT) en fonction du facteur de changement d'expression. **C.** L'union des gènes affectés dans les cellules TKO en termes d'expression ou d'épissage.

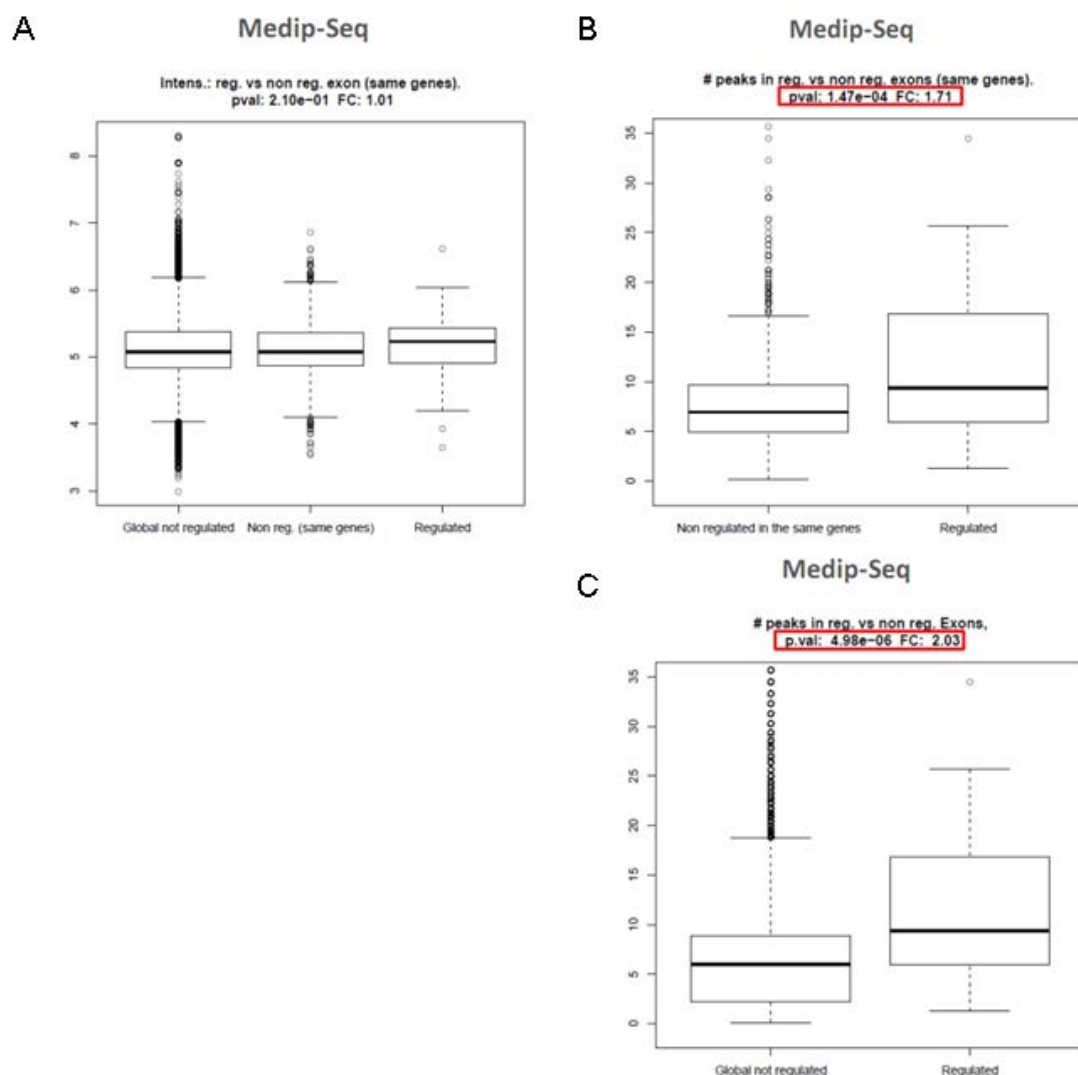


Figure 13. Les exons régulés par l'inactivation des DNMT sont enrichis en méthylation de l'ADN dans les cellules parentales. **A.** Intensité des pics de méthylation dans les exons régulés (regulated), les exons non régulés des gènes contenant un exon régulé (same genes) et l'ensemble des exons non régulés (Global not regulated). **B.** et **C.** Nombre de pics de méthylation dans les exons régulés (regulated) comparé aux exons non régulés des gènes contenant un exon régulé (same genes) (B) et l'ensemble des exons non régulés (Global not regulated) (C).

DISCUSSION ET PERSPECTIVES

1. L'INTERPRETATION DU COUPLAGE EPISSAGE - CHROMATINE EST COMPLEXE

1.1. Corrélation n'implique pas causalité

La plupart des études portant sur le couplage de l'épissage avec la chromatine sont majoritairement corrélatives et mènent parfois des sur-interprétations. En effet, de la seule constatation d'une corrélation entre deux phénomènes il n'est pas possible de déduire la nature de la relation qui les unit.

Beaucoup d'études sont basées sur l'observation à la suite de traitements de corrélations entre des modifications de décisions d'épissage et des modifications de marques épigénétiques au locus génique de l'exon alternatif. De ces corrélations, le modèle proposé repose sur le rôle de la chromatine au locus L sur l'épissage au locus L (**Figure 14**).

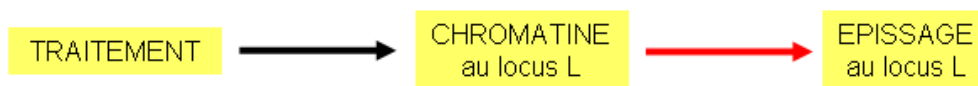


Figure 14. L'observation de corrélations entre des modifications de décisions d'épissage et des modifications de marques épigénétiques peut être interprétée comme une relation de cause à effet : le traitement utilisé affecte la chromatine au locus L ce qui engendre des changements de décisions d'épissage au locus L.

Les traitements utilisés dans ces études sont variés mais tous ont des conséquences globales sur la chromatine. Nous citons ci-dessous quelques exemples :

Traitement par des drogues inhibant des enzymes catalysant des marques chromatinienne :

- Le sodium butyrate (NaB), notamment utilisé dans Hnilicová *et al.* (Hnilicová et al., 2011) est un inhibiteur des HDAC (*Histone DeACetylase*). Il induit une augmentation globale de l'acétylation des histones H3 et H4 dans de nombreux types cellulaires de vertébrés (Candido et al., 1978). Par exemple, dans des cellules bovines l'analyse des données de ChIP suivie de séquençage à haut débit a mis en évidence que le traitement

au NaB provoque des enrichissements en acétylation des résidus H3K9 et H3K27 dans 10 000 et 50 000 régions respectivement (Shin et al., 2012).

- La trichostatine A (TSA) notamment utilisée dans Schor *et al.* (Schor et al., 2009) est aussi un inhibiteur des HDAC qui induit une augmentation globale de l'acétylation des histones sur de nombreux résidus (Lopez-Atalaya et al., 2013; Wang et al., 2009) notamment dans l'hippocampe (Lopez-Atalaya et al., 2013) région d'où proviennent les neurones dans les travaux de Schor *et al.*

Déplétion de facteurs chromatinienens par ARN interférence :

- La protéine HP1 γ d'après l'analyse des données de ChIP suivie de séquençage à haut débit dans 3 lignées cellulaires humaines a révélé, en prenant uniquement en compte les régions intragéniques, que cette protéine est présente dans environ 8000 gènes (Smallwood et al., 2012). Ainsi, la déplétion de HP1 γ notamment utilisée dans Saint-André et al. (Saint-André et al., 2011) est susceptible d'affecter des milliers de locus.

Ces quelques exemples mettent en exergue le fait que chaque traitement est susceptible d'affecter des milliers de locus dans le génome. L'interprétation des corrélations observées entre épissage et chromatine à un locus donné comme une relation de cause à effet est donc précoce. Dans toutes les études, la prise de recul par rapport au locus d'intérêt est nécessaire. La prise en compte de tous les locus chromatinienens affectés mettrait très certainement en évidence de nombreuses autres corrélations qui pourraient tout aussi bien expliquer les résultats obtenus. Par ailleurs, il n'est pas impossible que les différents traitements aient d'autres effets non liés à la chromatine (**Figure 15**).

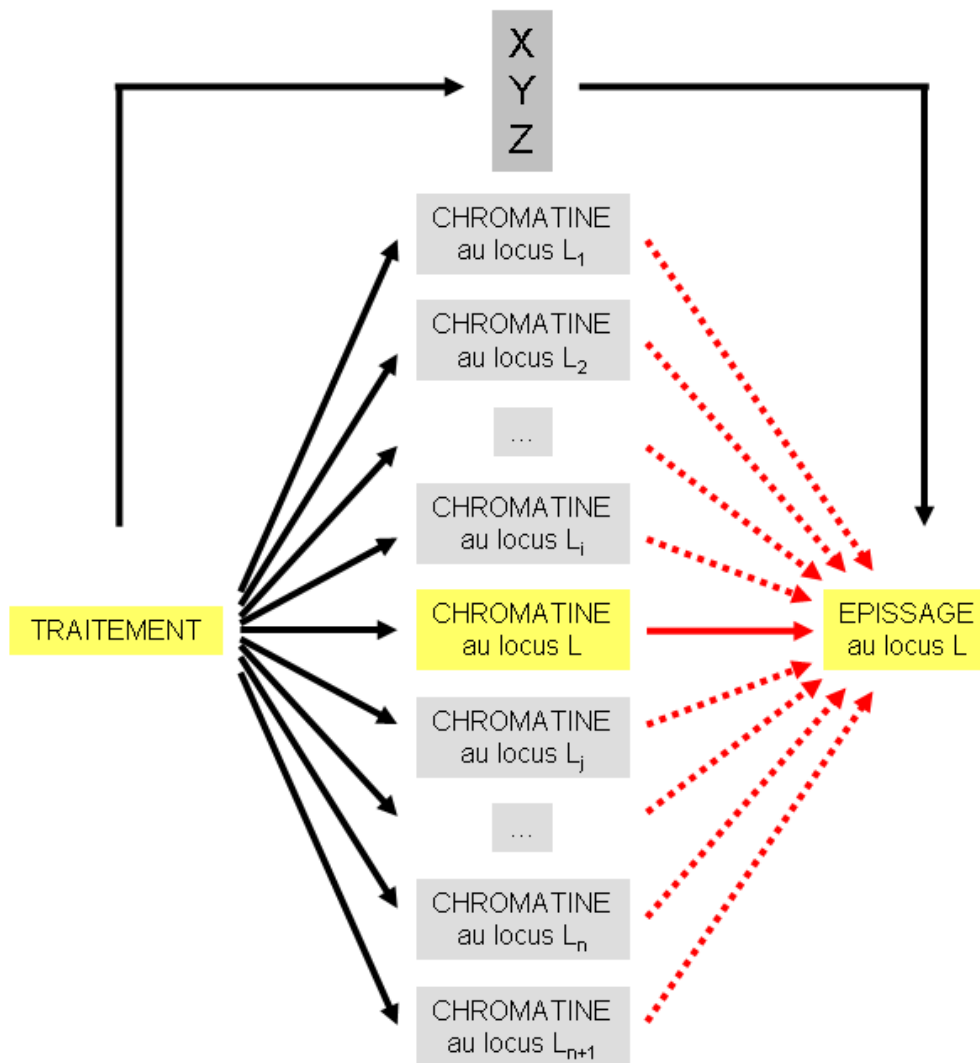


Figure 15. L'impact d'un traitement sur une décision d'épissage au locus L peut être interprété de nombreuses façons différentes. Le traitement affecte la chromatine à de nombreux loci L_i . Chacun de ces locis peut en théorie affecter l'épissage du locus L de manière directe ou indirecte (flèches pointillées). Le traitement peut aussi affecter l'épissage *via* des effets (X, Y, Z) non liés à la chromatine.

Nous allons à présent explorer d'autres interprétations possibles quant aux effets sur l'épissage d'un traitement ciblant la chromatine.

1.2. Les facteurs chromatinien peuvent réguler l'expression des facteurs d'épissage

De très nombreuses études ont mis en évidence l'implication des facteurs chromatinien et des marques épigénétiques dans la régulation transcriptionnelle. Ainsi, la déplétion de ces facteurs ou leur inhibition en utilisant des drogues ont des conséquences majeures sur l'expression des gènes. Dans les études du couplage épissage - chromatine, l'expression

différentielle de certains gènes notamment ceux impliqués dans la régulation de l'épissage, pourrait expliquer les différences d'épissage des gènes étudiés.

Traitement par des drogues inhibant des enzymes catalysant des marques chromatinienne :

- Une étude de Piotrowska *et al.*, a mis en évidence que les traitements à l'azadeoxycytidine, la trichostatine A et le butyrate de sodium augmentent l'expression de ASF/SF2 dans les cellules humaines HT-29 et MCF-7 (Piotrowska and Jagodzinski, 2009).
- Dans l'étude de Hnilicová *et al.*, les auteurs utilisent le traitement au butyrate de sodium qui inhibe les HDAC (Hnilicová et al., 2011). L'analyse du transcriptome met en évidence que le NaB induit des changements d'expression de Prp40, un composant du spliceosome, du facteur d'épissage SFSR10, et de plusieurs protéines de liaison à l'ARN. Pourtant, l'implication d'aucun de ces facteurs n'a été explorée.

Déplétion et invalidation de facteurs chromatinien par ARN interférence :

- Dans l'étude de Tolstorukov *et al.*, les auteurs suggèrent que H2A.Bbd est impliqué dans la formation d'une structure chromatinienne qui facilite l'épissage (Tolstorukov et al., 2012). Dans cette même étude, les auteurs montrent que la déplétion de ce variant d'histone provoque une dérégulation de l'expression de plus de 1200 gènes. Les gènes appartenant au groupe ontologique « processus métabolique des ARN » sont notamment enrichis en H2A.Bbd. L'hypothèse d'une régulation de l'expression de facteurs d'épissage n'a pourtant pas été évoquée.
- D'après l'analyse de nos données de séquençage à haut débit du transcriptome des cellules murines invalidées pour les 3 enzymes DNMT1, 3A et 3B, l'expression de Mbnl1 est augmentée près de 6 fois. Mbnl1 a été montré comme régulant l'épissage de nombreux gènes (Han et al., 2013).

Comme nous l'avons exposé précédemment, le spliceosome est une machinerie très dynamique et flexible ce qui lui confère un caractère hautement régulable. Ainsi, le changement d'expression de quelques facteurs d'épissage peut impacter sur l'assemblage du spliceosome ou la catalyse de la réaction d'épissage et par conséquent affecter l'épissage. Ainsi, les conséquences d'un traitement sur l'expression des facteurs d'épissage ne doivent pas être ignorées.

1.3. Les facteurs chromatinien peuvent réguler les modifications post-traductionnelles de facteurs d'épissage

Les composants du spliceosome et des facteurs d'épissage sont modifiés post-traductionnellement. Ces modifications jouent un rôle clef dans la progression de l'épissage. De nombreux exemples dans la littérature ont mis en exergue que les enzymes « chromatinien » modifient post-traductionnellement des substrats non-histones. La régulation des facteurs d'épissage par des modifications post-traductionnelles des facteurs « chromatinien » est une éventualité à considérer.

Phosphorylation / déphosphorylation :

Parmi les modifications post-traductionnelles, ce sont les plus étudiées. Les phosphorylations réversibles des composants du spliceosome ont des rôles décisifs dans l'assemblage du spliceosome et dans la catalyse de l'épissage. De manière très intéressante, certaines des enzymes catalysant les (dé)phosphorylations de facteurs impliqués dans l'épissage, modifient aussi post-traditionnellement les histones.

- **La kinase DYRK1A** a été caractérisée pour phosphoryler le composant de la snRNP U2, SF3b1 (de Graaf et al., 2006) et des protéines SR ASF/SF2 (Shi et al., 2008), 9G8 (Ding et al., 2012), SC35 (Qian et al., 2011). Au laboratoire, Suk Min Jang a montré que cette kinase catalyse la phosphorylation de la thréonine 45 et de la sérine 57 des histones H3 ; ces modifications interfèrent avec la fixation des protéines HP1 (article en revue dans *EMBO report*).
- **La phosphatase PP1** est capable de déphosphoryler la protéine U5-116kDa et la protéine associée à la snRNP U2, SF3b155 (Shi et al., 2006). Par ailleurs, PP1 est impliquée dans la balance des niveaux de phosphorylation des histones H3 (Hsu et al., 2000).

Méthylation / déméthylation :

- **L'arginine méthyl-transférase PRMT5** est une enzyme capable de méthyliser des protéines cytoplasmiques et nucléaires. Dans le cytoplasme, PRMT5 méthyle les protéines Sm, des composants du cœur des snRNP ; leur méthylation est requise pour un assemblage correct des snRNP (Bezzi et al., 2013; Friesen et al., 2001; Meister et al., 2001). Dans le noyau, PRMT5 est aussi connu pour méthyliser les arginines,

notamment H3R8 (Pal et al., 2003), H2AR3 (Ancelin et al., 2006) et H4R3 (Ancelin et al., 2006; Lacroix et al., 2008).

- **L'arginine méthyl-transférase PRMT1** cible aussi les histones et méthyle notamment les histones H4 au résidu R3 (Strahl et al., 2001). PRMT1 peut aussi méthyliser des substrats non-histones, notamment le facteur d'épissage à domaine SR : SRSF9/SRp30c (Bressan et al., 2009).
- **La protéine JMJD6** (*Jumonji domain-containing 6*) est impliquée dans la régulation de U2AF65 *via* une modification post-traductionnelle, la lysyl-5-hydroxylation (Webby et al., 2009). JMJD6 permet aussi la déméthylation des histones H3 et H4 aux résidus arginine 2 et arginine 3 respectivement (Chang et al., 2007).
- **L'arginine méthyl-transférase CARM1** catalyse la méthylation de plusieurs facteurs d'épissage tels que CA150, SAP49, SmB, et U1C (Cheng et al., 2007). CARM1 a aussi été caractérisée pour son action méthyl-transférase sur les histones H3 au résidu R17 (Ma et al., 2001).

Acétylation / déacétylation :

Des analyses protéomiques ont aussi révélés que de nombreux composants du spliceosome sont acétylés (Choudhary et al., 2009). De manière intéressante, certains inhibiteurs d'histones-acétylases et d'histones-déacétylases sont capables de bloquer l'épissage d'un ARNpré-m *in vitro* (Kuhn et al., 2009) montrant que ces enzymes peuvent affecter l'épissage indépendamment de la chromatine.

Ces quelques exemples illustrent qu'une enzyme cible souvent plusieurs substrats. Là encore, les études du couplage épissage - chromatine ne rendent pas compte de l'éventualité que le facteur chromatinien d'intérêt soit impliqué dans la régulation post-traductionnelle de facteur d'épissage indépendamment de la chromatine.

1.4. Autorégulation de l'épissage et de la chromatine

Quand bien même, les corrélations observées entre chromatine et épissage soient la conséquence d'une relation de cause à effets (au locus considéré), le sens de causalité n'est pas trivial : i) la chromatine affecte l'épissage, ii) l'épissage affecte la chromatine, iii) la chromatine et l'épissage s'auto-influence mutuellement (**Figure 16**).

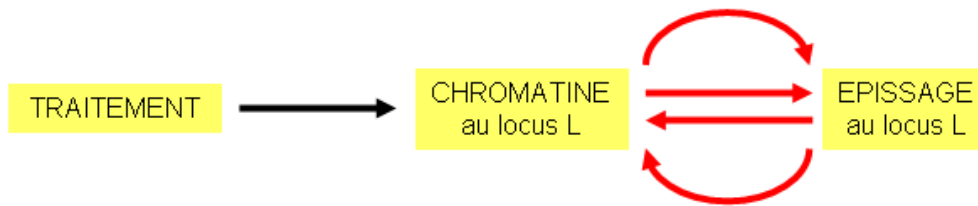


Figure 16. A un locus L donné, une relation de cause à effet entre épissage et chromatine peut s'interpréter des différentes manières.

Bien que corrélatives, certaines études suggèrent que la machinerie d'épissage régule les marques d'histones, notamment la méthylation de H3K36 *via* le recrutement de SETD2 (de Almeida et al., 2011). Dans une autre étude de Kim *et al.* utilisent un rapporteur du gène β -globine dont les sites d'épissage 3' sont mutés ou non intégrés à un site unique dans le génome. Les rapporteurs non compétant pour l'épissage présentent des niveaux plus faibles de méthylation en H3K36 (Kim et al., 2011a). Ainsi, ces résultats combinés avec ceux de Luco *et al.* et Pradeepa *et al.*, suggèrent que le couplage épissage - chromatine génère une boucle de rétrocontrôle. Pour étayer ces arguments, dans leur étude Zhou *et al.*, suggèrent que l'épissage de Nf1 et Fas est le résultat de l'autorégulation des protéines de liaisons à l'ARN Hu et des protéines HDAC (Zhou et al., 2011).

1.5. Bilan

Que la chromatine régule l'épissage de manière directe est une hypothèse à retenir mais d'autres modèles comme ceux que nous venons d'illustrer peuvent tout aussi bien expliquer les résultats. Par ailleurs, d'autres voies plus indirectes, telles que des effets sur le cycle cellulaire, l'apoptose... pourraient aussi rendre compte des résultats.

La difficulté de l'interprétation des études du couplage épissage - chromatine provient du fait que les approches expérimentales, c'est-à-dire l'utilisation de divers traitements ou la comparaison de différentes lignées cellulaires, provoquent de multiples effets à l'échelle de la cellule. Par conséquent, la distinction des effets directs et indirects de la chromatine sur l'épissage est impossible.

Ainsi, pour poursuivre l'étude sur le rôle de la méthylation de l'ADN sur la régulation de l'épissage alternatif, nous souhaitons développer des approches expérimentales qui permettront de limiter les paramètres variables.

2. DES APPROCHES EXPERIMENTALES PERMETTANT DE LIMITER LES PARAMETRES NON CONTROLABLES DOIVENT ETRE DEVELOPPEES

Pour déterminer si la méthylation de l'ADN affecte l'élongation transcriptionnelle, Lorincz *et al.* ont développé une approche ingénieuse : en utilisant un système Cre/loxP, les auteurs ont intégré à un locus spécifique un transgène méthylé ou non exclusivement dans la région intragénique (Lorincz et al., 2004). Ainsi, une approche similaire pourrait être utilisée pour déterminer si la méthylation intragénique affecte l'épissage alternatif de manière directe ; le seul paramètre variable serait l'état méthylé ou non du mini-gène.

Intégration du mini-gène dans le génome à un site unique :

Pour faire suite à nos travaux, il serait intéressant d'utiliser un mini-gène avec des exons alternatifs de CD44, v4-v5 par exemple. Des mini-gènes avec les exons cassettes EDI et EDII du gène de la fibronectine FN1 représenteraient aussi un intérêt particulier pour cette étude puisque leur épissage est sensible à l'azadeoxycytidine (Alló et al., 2009) (et nos données, non montrées).

Comme effectué dans l'étude de Lorincz *et al.* (Lorincz et al., 2004), cette approche nécessitera la méthylation *in vitro* de la région intragénique du mini-gène exclusivement. Cette étape peut par exemple se faire en utilisant l'enzyme SssI, une CpG méthylase (Matsuo et al., 1994).

Ensuite, le fragment correspondant à la région intragénique du mini-gène, méthylé ou non, sera ligué dans le vecteur recombinant (système Cre/lox, système FLP-In). Cette étape sera probablement la plus délicate, en effet, il faudra mettre au point les conditions expérimentales pour liguer suffisamment de matériel pour pouvoir transfecter le vecteur « fraîchement » ligué dans les cellules recombinantes. En effet, il n'est pas possible d'amplifier le vecteur dans des bactéries comme il est fait habituellement car la méthylation des CpG serait perdue. En effet, chez *E. Coli*, la méthylation des cytosines se fait sur la 2^{ème} cytosine des sites 5'CCWGG3' (Militello et al., 2012). Néanmoins, une alternative pourrait être testée : exprimer DNMT1 dans les bactéries et vérifier que son expression permet de maintenir la méthylation sur les CpG de la région intragénique du mini-gène sans que la méthylation ne se répande.

Vérification de l'état de méthylation après intégration dans le génome :

Un des risques de cette approche est que l'état de méthylation soit instable après

intégration dans le génome des cellules recombinantes. Il sera donc indispensable de vérifier que la méthylation reste au cours des réplifications et qu'elle ne se répande au niveau du promoteur. Dans Lorincz *et al.*, les auteurs ont montrés que 21 jours après intégration le niveau de méthylation est stable dans la région intragénique (Lorincz et al., 2004).

Evaluer l'épissage alternatif du mini-gène :

Préalablement, il faudra aussi s'assurer que le mini-gène dans sa version non méthylée exprime les différentes isoformes d'épissage. Ainsi, si la méthylation du mini-gène induit des différences d'épissage dans le sens de l'inclusion ou de l'exclusion les différences seront observables.

Ensuite, il s'agira de déterminer si l'épissage du mini-gène méthylé est différent. L'unique paramètre variable étant l'état de méthylation intragénique du mini-gène intégré, si les décisions d'épissage diffèrent, il s'agira bien d'un effet au locus du mini-gène et non des effets indirects *via* une expression différente de certains facteurs. Néanmoins, il sera essentiel d'utiliser plusieurs clones dans chaque condition pour s'assurer que les éventuelles différences ne soient pas dues à une adaptation différente des clones.

Perspectives :

Si des différences d'épissage sont observées cela ouvrira de très nombreuses questions quant aux mécanismes mis en jeu : ralentissement de l'ARN polymérase II, intervention des protéines MBD, positionnement des nucléosomes, modifications d'histones...

**PROJECT 2: ALTERNATIVE SPLICING EXTENDS
FUNCTIONS OF G9A AND SUV39H2 HISTONE
METHYLTRANSFERASES**

INTRODUCTION

Alternative splicing alters the coding sequence of mature mRNAs and is a source of proteome diversity. Transcriptome analysis has recently revealed that the impact of splicing on gene expression has been underestimated, and that almost every gene yields alternatively spliced transcripts (Pan et al., 2008; Wang et al., 2008). Most of these alternative-splicing events have not been characterized, and identifying their function represents a major challenge for biology.

Pre-mRNA splicing is a step-wise process catalyzed by the spliceosome, a macromolecular machinery containing five snRNPs and about hundred splicing factors (Jurica and Moore, 2003; Wahl et al., 2009). This enzymatic complex is highly dynamic allowing various flexible pathways to regulate alternative splicing. Although, the spliceosome does not strictly require other RNA processing machineries to remove introns (Krainer et al., 1984; Green et al., 1983; Ibrahim et al., 2005), it is now accepted that the transcription apparatus and epigenetic mechanisms influence splicing regulation *in vivo* (Dujardin et al., 2013; Luco et al., 2011; Braunschweig et al., 2013). In particular, the transcription machinery is involved in loading early splicing factors on nascent transcripts (Cramer et al., 1999; Das et al., 2007), while chromatin may directly or indirectly affect recruitment of splicing factors to sites of transcription (Batsché et al., 2006; Sims et al., 2007; Luco et al., 2010; Saint-André et al., 2011; Shukla et al., 2011; Pradeepa et al., 2012). Inversely, splicing factors can influence transcription (Lin et al., 2008; Ji et al., 2013) and affect the properties of chromatin (Schor et al., 2009; Zhou et al., 2011; De Almeida et al., 2011; Kim et al., 2011b; Keren-Shaul et al., 2013). These observations suggest that the crosstalk between splicing and chromatin machineries generates feed back loops controlling gene expression.

In parallel, alternative splicing engenders multiple isoforms of chromatin-remodeling and histone-modifying enzymes, thereby providing an additional pathway allowing splicing to affect chromatin structure and gene expression. The activity of the variant isoforms generated by alternative splicing is rarely approached in functional studies. Here, we have characterized the isoforms derived from two histone methyltransferases (HMTase), G9A (EHMT2) and SUV39H2. These enzymes belong to a family of six members, that also

include GLP (EHMT1), SETDB1, and SUV39H1 which cooperate to mono-, di-, or tri-methylate histone H3 at lysine 9 (H3K9me1/2/3) (Fritsch et al., 2010; Wu et al., 2010). These epigenetic marks are generally associated with transcriptional silencing and were recently shown to influence alternative splicing (Ameyar-Zazoua et al., 2012; Saint-André et al., 2011).

G9A is essential for embryonic stem cell differentiation and development. It specifically catalyzes H3K9me1/2 marks, enriched in euchromatin (Tachibana et al., 2001, 2002) while also detected in heterochromatic regions (Estève et al., 2005; Saint-André et al., 2011). The SUV39H1/H2 enzymes are H3K9 tri-methyltransferases, localized chiefly to pericentromeric regions (Aagaard et al., 1999, 2000; Carroll et al., 2000; Melcher et al., 2000). SUV39H2 was described originally as testis-specific (Carroll et al., 2000), but more recent observations from SUV39H1 and SUV39H1/H2 knock-out mice suggest that it may have functions in other tissues (Peters et al., 2001; García-Cao et al., 2004).

In the present study, we demonstrate that the *G9A* exon 10 and *SUV39H2* exon 3 are alternatively spliced, giving rise to two and three mature transcripts respectively. These transcripts are broadly expressed at different ratios in human tissues and cell lines. They all encode for proteins characterized by various technical approaches. Interestingly, the species conservation of *G9A* spliced variants and the lack of differences in their HMTase-related function, lead us to propose that exon 10 is associated to a non-histone function of this gene. Conversely, SUV39H2 isoforms bear differential HMTase activities, and their ratios are important to regulate gene expression. Therefore the regulation of chromatin factor expression by alternative splicing extends the crosstalk/relationships between splicing, transcription and chromatin. Finally, our data emphasize that gene expression and gene function are clearly not collinear, thus alternative splicing must be taken into account in any functional studies.

MATERIALS ET METHODS

Cell culture and siRNA transfections

Wi38 (ATCC), IMR90 (ATCC), Caco-2, HepG2, SKOV3 (HTB-77, ATCC), SW626 (HTB-78, ATCC), 293T cells were maintained in Dulbecco's modified Eagle's medium (Gibco, #31966-021) supplemented with 10% (v/v) foetal bovine serum (Thermo Scientific, #SV30160.03) and 100 U/mL penicillin-streptomycin (Gibco, #15140-122). HeLa (ATCC) cells were grown in the same medium supplemented with 7% (v/v) foetal bovine serum. Rhabdoid cells provide by O. Delattre's lab (MON, Medjkane *et al.*, Cancer Research, 2008) were maintained in Roswell Park Memorial Institute medium (Gibco, #61870-010) supplemented with 10% (v/v) foetal bovine serum and 100 U/mL penicillin-streptomycin. Cells were transfected with a mix containing siRNA (10-20 nM) and RNAi Max reagent (Life Technologies, #13778075) according to the reverse transfection protocol then collected 72 h after for analysis. For experiments in which the cells were transfected twice, the second round of transfection was performed 72 h after the first one. All siRNAs were synthesized as ON-TARGET*plus* grade from Dharmacon and are listed in **Supplementary Table S2**.

Tissues and organisms, RNA extraction, reverse transcription, radio-labelled and quantitative real-time PCR

Mouse, Chicken and Zebrafish samples were collected at indicated times and stored at -80°C during 4 months. Total RNA from human tissues were from Clontech (#636533 and #636643), and those from cells and organisms were extracted as described (Chomczynski & Sacchi, Nature Protocol, 2006) followed with DNase I (Roche, #04716728001) treatment. The quality of RNA was checked on agarose gel. The cDNA libraries were built using 0.5 µg RNA reverse transcribed with M-MLV reverse transcriptase (Life Technologies, #28025013) and either oligo dT (Thermo Scientific, #SO131), random hexamers (Sigma) or target specific primers (see Supplementary Table S2). Semi-quantitative PCR was performed with γ -32P 5'-end-labeled primers, and PCR products were verified by DNA sequencing. Quantitative real-time PCR (qPCR) was assayed in 10 µL reactions with Brilliant III Ultra Fast SYBR-Green QPCR Master Mix (Agilent, #600882). PCR reactions were carried out in a Stratagene MX3005p system with the following thermal profile: 5 min at 95°C, then 37 cycles of 10 sec at 95°C and 12 sec at 62°C qPCR assays were analyzed with an MxPro software as described

earlier (Batsché *et al.*, NSMB, 2006). The primers used for PCR are listed in **Supplementary Table S2**.

Western blot, antibodies

Total proteins were separated by electrophoresis on 4-12% gradient PAGE gels (Bio-Rad, #345-0124) and transferred on nitrocellulose membrane (Bio-Rad, #1620115) for western blot. The following antibodies were used for western blot: anti-G9A (Sigma, #HPA050550), anti-SUV39H2 (mix of 1/3 Active Motif, #61449 and 2/3 Proteintech, #11338-AP), anti-Brg1 (Euromedex, #2SN-2E12-AS), anti-V5 (Life Technologies, #R960-25), anti-tubulin (Abcam, #ab56676), anti-HP1 α (Euromedex, #2HP-2G9-AS), anti-H3 (Abcam, #ab1791), anti-H3K9me2 (Millipore, #07-441), anti-H3K9me3 (Abcam, #ab8898), anti-Rabbit IgG HRP (GE Healthcare, #NA934V), anti-Mouse IgG HRP (Abcam, #ab6808). Western blot signal was acquired on an Odyssey Fc system (Licor) and analysed with the Image Studio software.

Constructs and viral transduction

Various cDNA isoforms of Tomato, *G9A* and *SUV39H2* were cloned in pLVX-IRES-ZsGreen1 lentiviral vector (Clontech, #632187, see **Supplementary Figure S6A**) for overexpression in human and mouse cells of a bicistronic transcript encoding for the corresponding protein and the ZsGreen1 protein. Flag and V5 tag sequences were added upstream of each cDNA, and the SV40 NLS sequence was inserted in frame downstream of Tomato cDNA. Sequences of all these constructs were checked by sequencing and are available in **Supplementary Figure S6B**. For the virus production, 293T cells (500 cm² dish) at 50% confluence were transfected with a mix containing 80 μ g gag/pol vector (Addgene, #Plasmid 14887), 16 μ g pCMV-VSV-G vector (Addgene, #Plasmid 8454), 80 μ g pLVX construct and 352 μ g of PEI (MW ~25000; Polysciences Inc., # 23966); medium was changed after 6-8 h and cells were grown for 48h at 37°C. Medium were filtrated onto 0.45 μ m PVDF membrane (Millipore, SCHVU05RE) and virus particle were concentrated 200 times by centrifugation on sucrose cushion (20% sucrose, 10 mM Tris pH 7.5, 100 mM NaCl, 1 mM EDTA) at 28000 rpm (SW 28 rotor) for 1 h 30 min at 4°C. Concentrated viruses were aliquoted and stored at -80°C. Transduction efficiency of each virus in HeLa and 293T cells was estimated by detection of the ZsGreen1 fluorescence and experiments were performed 96 h post-transduction.

Immunofluorescence and antibodies

Transduced HeLa cells grown on coverslips were fixed 30 min with 3.7% paraformaldehyde (v/v) in phosphate-buffered saline (PBS), then permeabilized 5 min with triton X-100 0.2% (v/v) in PBS. The protein of interest was immuno-detected by incubation for 1 h with anti-V5 (Life Technologies, #R960-25, 1:200 dilution), or anti-H3K9me2 (Millipore, #07-441, 1:1000 dilution), or anti-H3K9me3 (Abcam, #ab8898, 1:1000 dilution), washed with PBS, incubated with Alexa Fluor 488-conjugated goat anti mouse or rabbit IgG antibodies (Life Technologies, #A-11029 or #A-11034, 1:1000 dilution) for 1 h, washed with PBS. The coverslips were mounted onto slides using antifade reagent (Life Technologies, #S36939) and immunofluorescence images were acquired on a Carl Zeiss Axio Observer Z1 microscope equipped with ApoTome module.

Protein stability assay

For protein stability assay, the proteins of interests were detected after treatment of HeLa cells with 2 μ M MG-132 (Merck Millipore, #474790), or with 20 μ M of cycloheximide (Sigma, #C-6255) at indicated times.

Cell fractionation

A 60 cm² plate of HeLa cells (90% confluence) was resuspended in 600 μ L of B1 buffer (10 mM Tris pH 7.9, 10 mM KCL, 1.5 mM MgCl₂, 0.34 M sucrose, 10% glycerol) complemented with 0.2% triton X-100 and incubated 5 min on ice. The supernatant F1 was harvested after centrifugation 1000 g, 1 min, 4°C and the packed nuclei washed once with B1 buffer before incubation in 600 μ L of B2 Buffer (3 mM EDTA / 0.2 mM EGTA) 30 min on ice. Again the supernatant F2 was collected after centrifugation at (1400 g, 1 min, 4°C), while the pellet was resuspended in 600 μ L of B3 buffer (25 mM Tris pH 7.4, 450 mM KCl, 2.5% glycerol, 0.3% NP-40) complemented with 12 μ L of Turbo DNase (Applied, #AM1907) and incubated 15 min at 37°C with gentle shaking. Thereafter, the extract was applied to sonication with a Bioruptor (Diagenode; 10 sec ON, 15 sec OFF, high intensity) and incubated again 15 min at 37°C; this step was repeated twice before centrifugation at 12000 g, 10 min, 4°C to collect the supernatant F3 and insoluble pellet. All buffers were complemented with 0.1 mM DTT, 10 μ M PMSF and protease inhibitor cocktail (Roche, #11873580001). The proteins separated by this procedure were analysed by western blot using a volume of each fraction corresponding to an equal number of cell.

Recombinant proteins and *in vitro* histone methyltransferase (HMT) assay

For each recombinant protein, two plates (150 cm²) of 293T cells were transduced with virus issued from the corresponding pLVX construct. After 48 h, the cells were collected and fractionated as described above. The F1, F2, F3 fractions were pooled together and filtered onto 0.45 µM cellulose acetate membrane (Costar, #8163). Extracts were incubated with 20 µL of packed anti-Flag M2 agarose (Sigma, #A2220) 3 h at 4°C under rotation, then washed two times with HS buffer (25 mM Tris pH 7.4, 450 mM KCl, 2.5% glycerol, 0.3% NP-40) and once with the LS buffer (25 mM Tris pH 7.4, 150 mM KCl, 2.5% glycerol, 0.3% NP-40). Finally, the isolated proteins were eluted twice with 50 µL of 100 µg/mL 3X FLAG peptides (Sigma, #F4799) with shaking, 15 min at room temperature. Purified proteins were stored at -80°C in LS buffer supplemented with 10% glycerol, and their quantity estimated by western blot using anti-V5 antibody. For *in vitro* HMT assay, the purified Flag-V5-tagged proteins and 1 µg of recombinant histone H3.1 (NEB, #M2503S) were incubated in HMT buffer (1 mM of S-adenosylmethionine (NEB, #B9003S), 25 mM Tris pH 8, 10% glycerol) 1 h at 37°C. Then, *in vitro* reaction were analysed by western blot to follow histone methylation levels as indicated in figure.

High-throughput RNA sequencing and bioinformatics analysis

HeLa cells were transduced during 96 h with the corresponding viruses, and total RNA extracted as described above. The RNA quality was evaluated by capillary electrophoresis (Agilent 2100 Bioanalyzer system) then cDNA libraries were prepared, split and ligated with adaptators. Single end (1x50bp) sequencing was performed using Illumina HiSeq 2500 technology and each sample was sequenced in triplicate with a minimum of 25 millions sequenced reads per replicate. Details on number of sequenced reads for each sample are given in Supplementary Table S3. The RNA-Seq data analysis was performed by GenoSplice (<http://www.genosplice.com>). First, reads were aligned onto the human genome (hg19) using Bowtie software (v.0.12.7). Unmapped reads were aligned on an exon-exon junction database built using annotations from version 2013_1 of FAST-DB (see <http://www.easana.com>). This exon-exon junction database was constituted of known junctions, and putative junctions coming from the combination of all possible exon boundaries. For each gene present in FAST DB v2013_1, reads aligning on constitutive regions (that are not prone to alternative splicing) were counted. Based on these read counts, normalization and differential gene expression were performed using DESeq (v1.12.0 on R v3.0.0). EASANA was used for visualization of results (<http://www.easana.com>).

Chromatin ImmunoPrecipitation (ChIP) experiments

The cells were fixed with 1% (v/v) formaldehyde for 10 min at room temperature and the chromatin preparation was processed as described (Batsché *et al.*, NSMB, 2006). Chromatin was incubated overnight with 1 µg of anti-H3K9me3 (Abcam, #ab8898), or anti-H3 (Abcam, #1791), or non immune IgG antibodies (used as control) and saturated magnetic beads coupled to anti-rabbit antibodies (Dynabeads, Invitrogen, #11204D) were incubated during 2 h to recover the complexes. The beads were washed extensively and nucleic acids were purified and quantified by qPCR.

RESULTS

In silico analysis of approximately 100 different genes encoding chromatin factor revealed that a majority generates multiple transcripts, mainly as a consequence of alternative splicing (data not shown). These transcripts are potentially translated into distinct proteins, which lead to a diversity that is mostly underestimated in functional studies.

Therefore, we have examined here the functional issues of the various protein isoforms encoded by two HMTs, namely *G9A* and *SUV39H2*. More specifically, we have focused our study on alternative splicing of *G9A* cassette exon 10 that can be either skipped (*G9A_Δe10*) or included (*G9A_e10*), and on alternative splicing of *SUV39H2* exon 3 that can be either skipped (*SUV39H2_Δe3*), fully (*SUV39H2_e3L*) or partially (*SUV39H2_e3S*) included due to the presence of an internal alternative 5' splice site (**Supplementary Figure S1A**).

1. ALTERNATIVELY SPLICING AFFECTS A SIGNIFICANT PART OF *G9A* AND *SUV39H2* TRANSCRIPTS

To evaluate first, the transcriptional levels of both HMTases, total RNAs from 21 human tissues were quantified by RT-qPCR using primers detecting a constant region of the genes (**Supplementary Figure S1B**). Levels of these total RNAs were controlled by housekeeping genes expression that exhibited a constant expression among tissues (**Figure 1A, see RLP0**). Thus, we showed that the *G9A* mRNA is broadly expressed and as previously reported *SUV39H2* mRNA is highly expressed in testis (O'Carroll et al., MCB, 2000) but definitively not restricted to this tissue.

Therefore, we analyzed these human tissues at the level of alternative splicing decisions by semi-quantitative RT-PCR using primers framing the variant exons of *G9A* and *SUV39H2* (**Figure 1A**). This revealed that alternative splicing of *G9A* exon 10 and *SUV39H2* exon 3 are frequent and ubiquitously detectable in every tissue, suggesting that this regulation pathway is important to regulate the total pool of mRNA expressed by respective genes (**Figure 1A**).

Proportions of skipping *versus* inclusion are highly variable from one tissue to the other and the inclusion of variant exon is independent of the global transcriptional levels of the genes. Surprisingly, we also found that the full-length isoform is often not the more abundant. In fact, in tissues such as kidney, thymus and testis, *G9A_Δe10* is preponderant. Likewise, shortened e3 inclusion isoform *SUV39H2_e3S* is the main transcript in cerebellum, fetal liver,

lung, skeletal muscle, thymus, stomach and testis, and the skipped isoform *SUV39H2_Δe3* is preponderant in heart, liver, salivary gland, spleen, thyroid gland, trachea and uterus. In some of these tissues, full-length *SUV39H2_e3L* transcripts are minor almost undetectable species (heart, liver, and lung). Additional splice *SUV39H2* variants may also be generated as suggested by uncharacterized species detected in brain and adrenal gland and prostate (**Figure 1A**).

Alternative splicing regulation of both HTMs was also observed in eight different cell lines with either an epithelial (ovarian SW626, colorectal Caco2, and hepatic HepG2 cells), a mesenchymal (pulmonary fibroblasts IMR90 and Wi38), or a dedifferentiated (ovarian SKOV3, cervical HeLa, and rhabdoid MON cells) origins, with an interesting accumulation of *G9A_Δe10* in the epithelial cells (**Supplementary Figure S1C, D, E**).

Together, these observations indicate that previous functional study and gene expression dosage performed with only one transcript are clearly not sufficient to appreciate *G9A* and *SUV39H2* gene function.

To confirm that these alternative transcripts result really of a dynamic modulation by alternative splicing and not of the leaking expression issued of a pseudogene for instance, we used siRNAs to individually deplete about 50 different splicing factors in either MCF7, or SKOV-3ip1 cells. This approach showed that the depletion of more than half of the tested factors affects splicing of *G9A* exon 10 and *SUV39H2* exon 3 (**Supplementary Figure S1F**). Moreover, 17 and 18 splicing regulators modulate splicing of *G9A* exon 10 and *SUV39H2* exon 3 in the two different cell lines suggesting that these factors could act directly on *G9A* and *SUV39H2* splicing. Notably, the RBM9/RBFOX2 which regulates the *SUV39H2* exon 3 splicing (**Figure 1B**) was shown to interact this exon by iCLIP-seq in human embryonic stem cells (Yeo et al. 2009) (**Supplementary Figure S1G**). We concluded that both exons are highly controlled by alternative splicing.

Moreover, *G9A* exon 10 is conserved in mouse and also alternatively spliced in E12.5 embryos (**Figure 1C**). *G9A* gene is absent in chicken and the locus containing the ortholog *G9A* exon 10 is not conserved in Zebrafish. *SUV39H2* exon 3 is conserved in ortholog genes and alternatively spliced in mouse, zebrafish (with changes between embryonic and adult stages), and in chicken (with changes between embryos at day 1 (D1) and day 7 (D7)) (**Figure 1C**). Therefore, we concluded that the splicing regulation of both alternative exons is

conserved among species suggesting that it plays an important role in the function of two genes.

All these observations showed that a significant part of HMTase transcription results in various alternative spliced isoforms suggesting that they could play an important role in the gene function of these HMTases.

2. ALTERNATIVE SPLICED TRANSCRIPTS OF G9A AND SUV39H2 ENGENDER VARIANT PROTEINS

Skipping of *G9A_e10* results in an in-frame deletion in the mRNA and is predicted to generate a version of G9A shorter by 33 amino acids (**Figure 2A**). Consistent with this, western blotting with an anti-G9A antibody on extract from HeLa cells revealed two G9A species with approximatively the same gel mobility, both affected by depletion of *G9A* mRNA isoforms with siRNAs (**Figure 2B, 2C**). These species represented the main G9A isoforms in a series of cell lines (**Supplementary Figure S2A**). The mobility of these species corresponds approximatively with the predicted molecular weight of G9A_e10 and G9A_Δe10 proteins (132 and 129 kDa, respectively) and their respective accumulation correlate with that of a *G9A_e10* and *G9A_Δe10* transcripts (**compare Supplementary Figure S2A to Supplementary Figure S1C**). Together, these observations indicate that alternative splicing of *G9A_e10* and *G9A_Δe10* pre-mRNAs results in the production of two G9A protein isoforms.

The region deleted in G9A_Δe10 is an intrinsically disordered region located immediately downstream of the G9A nuclear localization signal (NLS) and upstream of a series of ankyrin (ANK) repeats, involved in protein-protein interactions, and the catalytic SET domain (**Figure 2A**). Such disordered regions are frequently encoded by alternative exons and allow for increased conformational heterogeneity, provide additional surfaces for protein-protein interactions, or contain sites potentially modified post-translationally (Buljan et al., 2013; Ellis et al., 2012). The region encoded by *G9A* exon 10 contains 21% serines or threonines, but analysis of the databases revealed no direct evidence for the presence of post-translational modification at these residues.

Both full and partial skipping of *SUV39H2* exon 3, while maintaining the open reading frame in the *SUV39H2* mRNA, greatly affect the structure of SUV39H2 protein. SUV39H2_e3S lacks an intact catalytic SET domain, while SUV39H2_Δe3 is in addition missing the chromodomain involved in interaction with methylated H3K9 (**Figure 2A**). Western blotting analysis of whole cell extract with anti-SUV39H2 antibodies revealed 3 species (**Figure 2C and Supplementary Figure S2A**).

The estimated sizes of these species are compatible with them being the predicted product of SUV39H2_e3L, SUV39H2_e3S, and SUV39H2_Δe3 alternatively spliced mRNAs respectively (49, 29 and 24 kDa). The apparent molecular weight of SUV39H2_e3L is consistent with those showed by O'Carroll *et al.* (Carroll et al., 2000). Yet, levels of accumulation of the different species did not correlate well with those observed for the respective mRNAs (**compare Supplementary Figure S2A to Supplementary Figure S1C right**). For instance, in IMR90 and WI38 where *SUV39H2_e3L* transcripts are preponderant, the longest SUV39H2 protein isoform is absent. Likewise, *SUV39H2_e3S* transcripts were abundant in most cell lines examined, while the medium sized protein isoform was in each case almost undetectable (**compare Supplementary Figure S1C_right to Supplementary Figure S2A**).

Depletion of *SUV39H2* mRNA with siRNAs targeting an exon-exon junction present in all the splice variants allowed us to verify that the different species were all *SUV39H2* gene products (**Figure 2C**). These experiments also revealed that reduced accumulation of the larger SUV39H2 protein isoform was reached only 5 days after transfection of the siRNAs, while the smaller isoforms became undetectable after only 3 days (**Supplementary Figure S2B**).

These observations indicate that *SUV39H2_e3L*, *e3S* and *Δe3* transcripts are all encoded into proteins. Differences in the stability of the various SUV39H2 protein isoforms may explain why their levels of accumulation cannot be predicted from the levels of mRNA.

3. ALTERNATIVE SPLICING REGULATES STABILITY OF G9A AND SUV39H2 PROTEINS

To further explore the variations in stability observed for the SUV39H2 isoforms and to gain insight into eventual other differences that may characterize each G9A and SUV39H2 isoforms, we used a lentivirus-based expression system to expressed FLAG-V5-tagged (FV)

versions of these isoforms: FV_{G9A_e10} , $FV_{G9A_Δe10}$, $FV_{SUV39H2_e3L}$, $FV_{SUV39H2_e3S}$, $FV_{SUV39H2_Δe3}$. This system also allowed for the bicistronic expression of ZsGFP (**Supplementary Figure S6A, S6B**). Gel mobility of these different isoforms was very similar to that observed for the endogenous proteins detected in cell extract when adjusted for the size of the affinity tag (**Figure 3A top**). This shows also that antibodies recognizing endogenous proteins are insensitive to eventual epitope masking. As observed for the endogenous SUV39H2 transcripts, the correlation between levels of each transcript variant with levels of the cognate proteins was poor, especially for $FV_{SUV39H2_e3S}$ (**compare Figure 3A to Supplementary Figure S3A**).

Treatment of transduced cells with cycloheximide (CHX), a protein synthesis inhibitor, showed that the two FV_{G9A} isoforms had comparable half-life of (1 to 2 h – **Figure 3B and Supplementary Figure S3B**), while $FV_{SUV39H2_e3L}$ was significantly more stable than the $FV_{SUV39H2_e3S}$ and $FV_{SUV39H2_Δe3}$ isoforms (more than 12 h *versus* less than 1 h). The relatively long half-life of $FV_{SUV39H2_e3L}$ is consistent with the slow decrease of the full length endogenous protein after knocking-down (**Supplementary Figure S2B**). Inhibition of the proteasome with MG132 specifically resulted in increased accumulation of $FV_{SUV39H2_e3S}$ and $FV_{SUV39H2_Δe3}$, confirming that these isoforms are subject to rapid turnover. Otherwise, MG132 treatment doesn't affect FV_{G9A} isoforms accumulation (**Figure 3C, Supplementary Figure S3B**), indicating that the variant exon-10 do not serve to tag the G9A protein for proteasome-mediated regulation.

4. ALTERNATIVE SPLICING REGULATES SUB-NUCLEAR LOCALIZATION OF SUV39H2 PROTEINS

Immunofluorescent microscopy in transduced HeLa cells showed that the two FV_{G9A} isoforms had both a nuclear diffused distribution with exclusion from the nucleoli (**Figure 3D**). $FV_{SUV39H2_e3L}$ displayed a very similar distribution while $FV_{SUV39H2_e3S}$ and $FV_{SUV39H2_Δe3}$ were unexpectedly distributed in foci (**Figure 3D**). This observation led us to investigate whether the G9A and SUV39H2 isoforms were affected in their ability to interact with chromatin. To that end, we used a previously described technique that divides cellular components into a detergent-extractable fraction of proteins either cytoplasmic, nucleoplasmic or loosely attached to chromatin (**Figure 3F, Lines F1 and F2**), a chromatin fraction (**enriched in histones, see Figure 3F, lines F3 and P**) accessible to DNase and

containing most of the RNAPII, and finally a fraction of insoluble chromatin largely depleted of RNAPII (Saint-André, 2011) (**Figure 3E**).

The two $_{FV}G9A$ isoforms showed similar extractability, and were present in both the detergent-extractable fraction (predominantly) and in the DNase-accessible chromatin fraction (**Figure 3F**).

In contrast, $_{FV}SUV39H2$ isoforms were found in different fractions. Full length $_{FV}SUV39H2_e3L$ was present almost exclusively in the DNase-accessible chromatin fraction, while $_{FV}SUV39H2_e3S$ and $_{FV}SUV39H2_Δe3$ were highly enriched in the detergent-extractable fraction (**Figure 3F**). Endogenous $Suv39H2_e3L$ and $SUV39H2_Δe3$ isoforms segregated in a similar manner than tagged-protein isoforms (**Figure 3F, bottom**).

These observations show that alternative splicing regulation leads to change of localization and stability of SUV39H2 proteins whereas the two G9A isoforms exhibit an equivalent pattern.

5. PART OF SUV39H2 HMTASE DOES NOT CARRY ENZYMATIC ACTIVITY

We next examine the catalytic activity of the G9A and SUV39H2 isoforms on the H3K9 methylation.

In an *in vitro* HMTase assay, immunopurified $_{FV}G9A_e10$, $_{FV}G9A_Δe10$, and $_{FV}SUV39H2_e3L$ showed comparable abilities to methylate recombinant histone H3 at K9 (**Figure 4A**). As expected, $_{FV}SUV39H2_e3S$ and $_{FV}SUV39H2_Δe3$ that lack an intact SET domain had no HMTase activity in this assay (**Figure 4A**). Interestingly, in transduced HeLa cells, $_{FV}SUV39H2_e3L$, but not $_{FV}G9A$, $_{FV}G9A_Δe10$, caused global increase in levels of H3K9me3 as observed by immunofluorescent microscopy (**Figure 4B, Supplementary Figure 4A**). This increase in levels of H3K9me3 induced by $_{FV}SUV39H2_e3L$ was also visualized in western blots, concomitant with a drop in levels of H3K9me2 (**Figure 4C**), observable also by immunofluorescence (**Supplementary Figure S4B**).

Together, these data suggest that the full length SUV39H2 is strongly tethered to chromatin and that its levels of accumulation directly determine the balance between cellular levels of H3K9me2 and H3K9me3. This activity is regulated by alternative splicing that affects both the sub-nuclear localization and the catalytic activity of the SUV39H2 proteins. In contrast, enforced expression of G9A isoforms that all retain catalytic activity but that are

poorly associated with chromatin does not affect levels of H3K9me3, possibly because G9A isoforms require molecular partners to be properly targeted to their substrate *in vivo*.

6. SUV39H2 ISOFORMS HAVE DIFFERENT TARGET GENES

We next investigated the activity of the G9A and SUV39H2 isoforms in transcriptional regulation. To that end, we forced expression of these isoforms in HeLa cells using the lentiviral constructs and examined transcription of *H19*, a gene previously reported as regulated by H3K9 methylation (Monnier et al., 2013; Nativio et al., 2011). In these experiments, all isoforms with an intact catalytic SET domain, including the two G9A isoforms and SUV39H2_e3L, functioned as transcriptional repressors, while SUV39H2_e3S and SUV39H2_Δe3 had a moderate activating effect on *H19* (**Supplementary Figure 5A**). Chromatin immunoprecipitation (ChIP) assays confirmed that only SUV39H2_e3L, but not SUV39H2_e3S and SUV39H2_Δe3 induced increased accumulation of H3K9me3 at the *H19* promoter (**Supplementary Figure 5B**).

To identify novel genes regulated SUV39H2_e3L and to possibly identify genes regulated by SUV39H2_e3S, we performed high throughput sequencing of mRNAs isolated from HeLa cells transduced to express either SUV39H2_e3L or SUV39H2_e3S. Cells transduced with an empty pLVX vector were used as a negative control. Forced expression of SUV39H2_e3L resulted in the up-regulation of 230 genes, while only 45 were down regulated (**Figure 5A**, **Supplementary Figure 5C left**).

To validate part of these data, we confirmed in independent RT-qPCR experiments that the *CNN1* gene was down-regulated by increased levels of SUV39H2_e3L. The shorter SUV39H2 isoforms did not cause repression at this gene (**Supplementary Figure 5D**). ChIP assays further confirmed that the repressive effect of SUV39H2_e3L was accompanied by increased accumulation of H3K9 trimethylation at the promoter of the *CNN1* gene (**Supplementary Figure 5B**).

We further confirmed that SUV39H2_e3L up-regulated a series of genes including *CD2*, *DHRS2*, *ACTL8*, *MAGB6* and *LINC1112* (**Figure 5C**). This up-regulation was abrogated upon subsequent depletion of SUV39H2 transcripts with siRNAs. In these experiments, shorter SUV39H2 isoforms were considerably less efficient at activating the genes under scrutiny. Surprisingly, ChIP analysis of the promoter of one of these genes (*CD2*) revealed that activation by SUV39H2_e3L is associated with increased levels of H3K9me3

(Supplementary Figure 5B). While we can not exclude that a non annotated alternative promoter could drive the CD2 transcription, this suggests that the effect of SUV39H2_e3L on this gene is direct and more generally illustrates that H3K9me3 is not systematically associated with transcriptional repression.

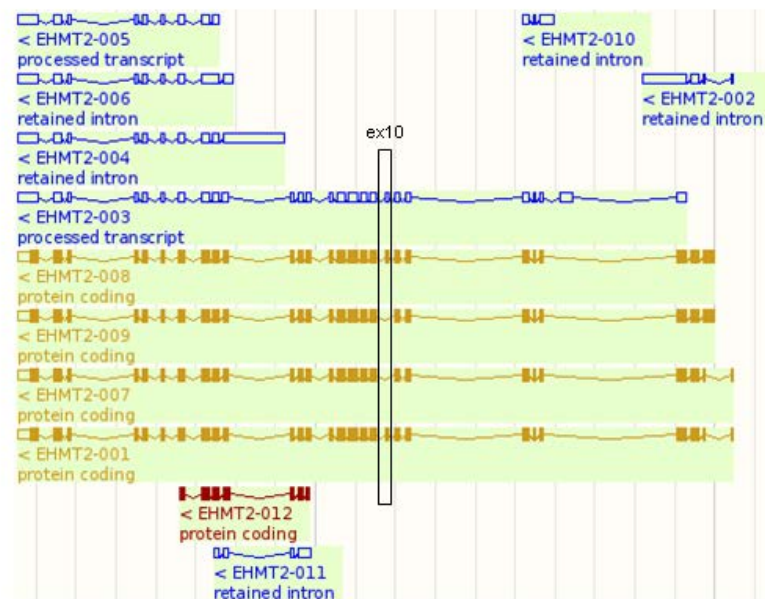
Genome-wide transcription analysis in the presence of elevated levels of SUV39H2_e3S resulted, like observed for SUV39H2_e3L, in more activated than repressed genes (271 genes up *versus* 47 down) **(Figure 5A and Supplementary Figure 5C).**

Yet, a fraction of these genes were affected only by SUV39H2_e3S, and not SUV39H2_e3L (275 out of the total 318 genes), suggesting that the two isoforms regulate a distinct set of genes. Consistent with this, we noted that 77 genes were differentially regulated by SUV39H2_e3L and SUV39H2_e3S: 49 genes are decreased and 28 are increased by e3S compared to e3L **(Figure 5B, list in Supplementary Table 1).** We finally used the pool of genes regulated by SUV39H2_e3S and SUV39H2_e3L to screen for genes regulated by the third SUV39H2_Δe3 isoform. This approach pinpointed *CCL5* as activated more efficiently by SUV39H2_Δe3 than by SUV39H2_e3L or SUV39H2_e3S **(Supplementary Figure 5E).** This activation was lost upon depletion of the SUV39H2 isoforms with siRNAs. This shows that SUV39H2_Δe3 may have a transcriptional regulator function and that the regulation of SUV39H2 expression by alternative splicing is not just a way to regulate the level of the catalytic active isoform.

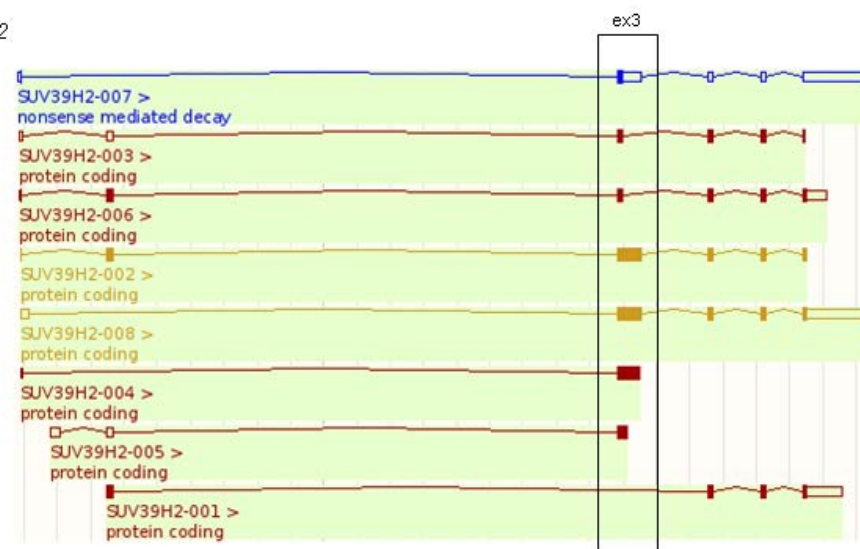
To check that endogenous SUV39H2 isoforms also regulate the genes identified using the SUV39H2 transgenes, we took advantage of the two human cell lines Wi38 and SW626 that each express a single SUV39H2 isoform **(Figure 5D) (SUV39H2_e3L and SUV39H2_Δe3 respectively – Supplementary Figure S2A, lanes 2 and 3).** Consistent with our observations in transduced HeLa cells, *CNN1* was expressed at a 10-fold lower level in the SUV39H2_e3L-expressing SW626 cells when compared to the Wi38 cells that lack this isoform **(Supplementary Figure 5F).** Likewise, the *DHRS2*, *ACTL8*, *MAGB6*, and *LINC1112* genes that were activated by transduced SUV39H2_e3L were also expressed at highest levels in SW626 cells. Neither of the two cell lines expressed the *CD2* gene **(Figure 5E).**

Together, these data define SUV39H2_e3S, SUV39H2_e3L, and SUV39H2_Δe3 as transcriptional regulators that function mainly as activators with each a specific subset of target genes.

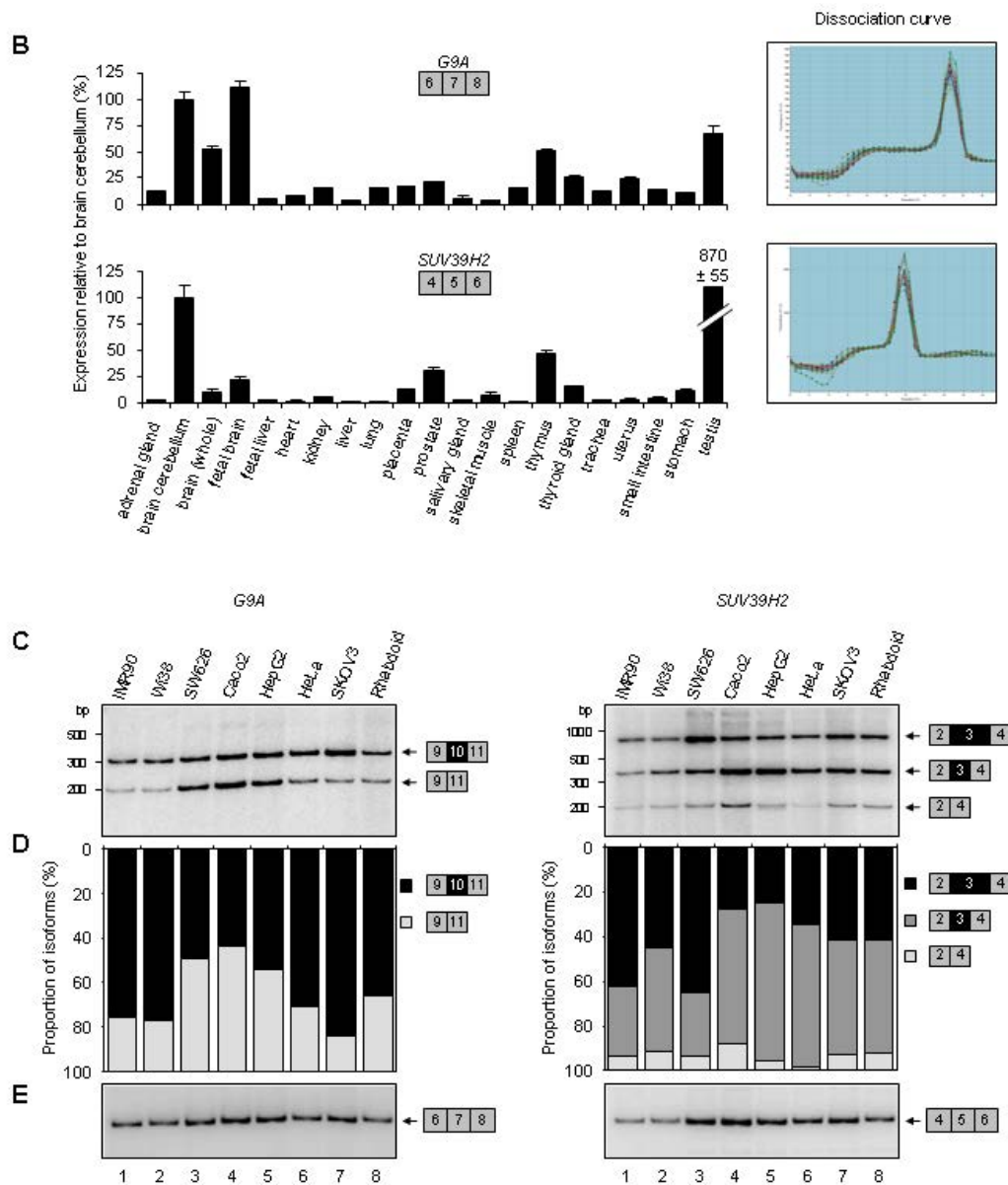
A
G9A



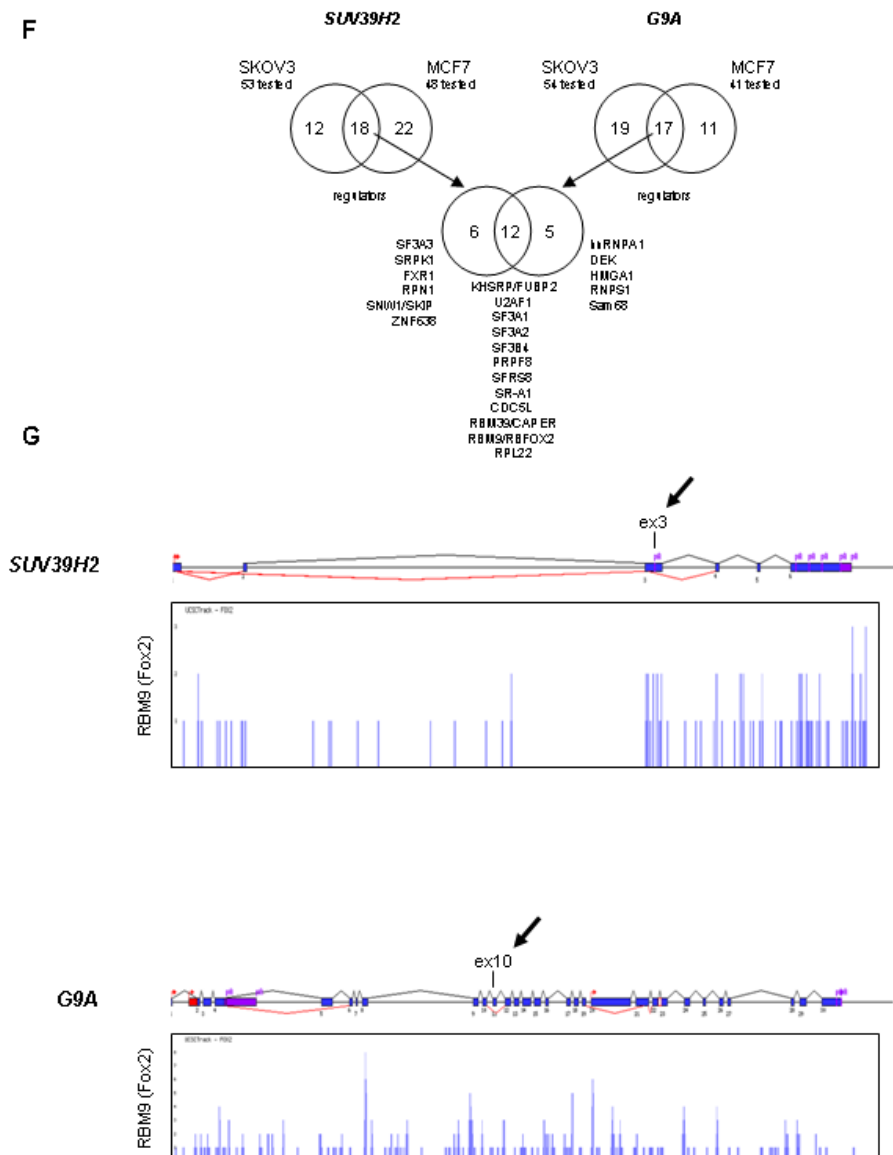
SUV39H2



Supplementary figure S1



Supplementary figure S1



Supplementary figure S1

Supplementary Figure S1. Alternative splicing regulates the expression of *G9A* and *SUV39H2* transcripts. (A) Screenshot displaying *G9A* (top) and *SUV39H2* (bottom) transcripts annotated in *ensembl* (www.ensembl.org). (B) Analysis in 21 human tissues of *G9A* and *SUV39H2* total expression relative to brain cerebellum. Primers used are localized in constant exons, dissociation curves of amplicons (right panels) confirm that there is only one PCR product. (C) and (D) Analysis in 8 human cell types of exon 10 inclusion into *G9A* transcripts and exon 3 inclusion into *SUV39H2* transcripts. (C) Semi-quantitative RT-PCR was performed with radiolabeled primers and total RNA extracted from the indicated cell types. (D) Proportions of each isoforms were determined quantifying radioactive signal. (E) Total *G9A* and *SUV39H2* expressions were analysed performing semi-quantitative RT-PCR with radiolabeled primers localized in constant exons. (F) Venn diagram displaying number of splicing factors affecting splicing of *G9A* exon 10 and *SUV39H2* exon 3 in SKOV and MCF7 cells. (G) Screenshot from *fasterdb* (<https://fasterdb.lyon.unicancer.fr/>) displaying RBM9 associated read number sequenced along *G9A* and *SUV39H2* transcripts in human embryonic tem cells according to CLIP-seq data (Yeo *et al.*, NSMB, 2009).

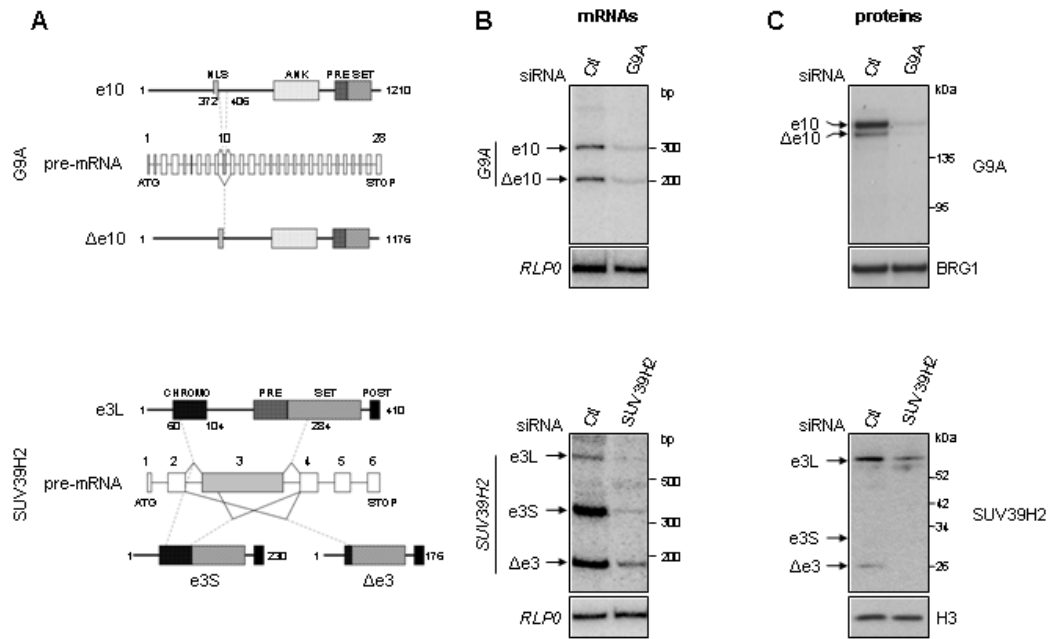
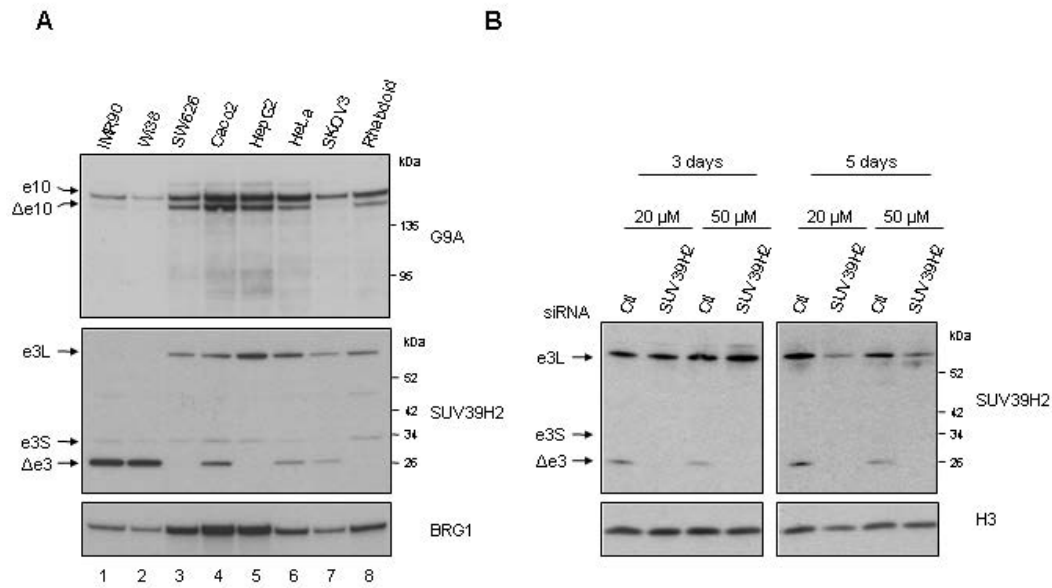


Figure 2

Figure 2. Alternatively spliced transcripts of G9A and SUV39H2 encode proteins. (A) Schematic view displaying the protein isoforms encoded by the transcripts regulated by alternative splicing of *G9A* exon 10 and *SUV39H2* exon 3. For pre-mRNA, introns are drawn with bold lines (—), exons with boxes (□), alternative exons of interest with gray boxes (■), alternative splicing events of pre-mRNA are drawn with sharp lines (∧) and the correspondence on protein isoforms are indicated with dotted line (⋯). Known protein domains are reported as follows: nuclear localization signal (NLS), ankyrin (ANK), pre-SET (PRE), SET (SET), post-SET (POST), chromodomain (CHROMO). (B) and (C) Analysis in HeLa cells of mRNAs and proteins resulting from alternatively spliced exon 10 and 3 of *G9A* and *SUV39H2* transcripts. The specificity of signals detected by RT-PCR (B) or Western blot (C) was checked by the comparative analysis of HeLa cells transfected or not with a siRNA targeting all *G9A* or *SUV39H2* transcripts. *G9A* and *SUV39H2* transcript levels were assessed 3 days after transfection and proteins levels were assessed after 3 days for *G9A* et 5 days for *SUV39H2*. Detection of *RLP0* transcripts and BRG1 or H3 proteins were used as loading controls.



Supplementary figure S2

Supplementary Figure S2. Alternative spliced transcripts of G9A and SUV39H2 encode for protein factors. (A) Analysis in 8 human cells types of proteins resulting from alternative spliced exon 10 and 3 of *G9A* and *SUV39H2* transcripts by Western Blot using anti-G9A and anti-SUV39H2 antibodies. Detection of BRG1 was used as loading controls. (B) Analysis of SUV39H2 protein depletion in HeLa cells transfected (*SUV39H2*) or not (Ctl) with 20 or 50μM siRNA targeting *SUV39H2* transcripts during 3 or 5 days. Detection of H3 was used as loading controls.

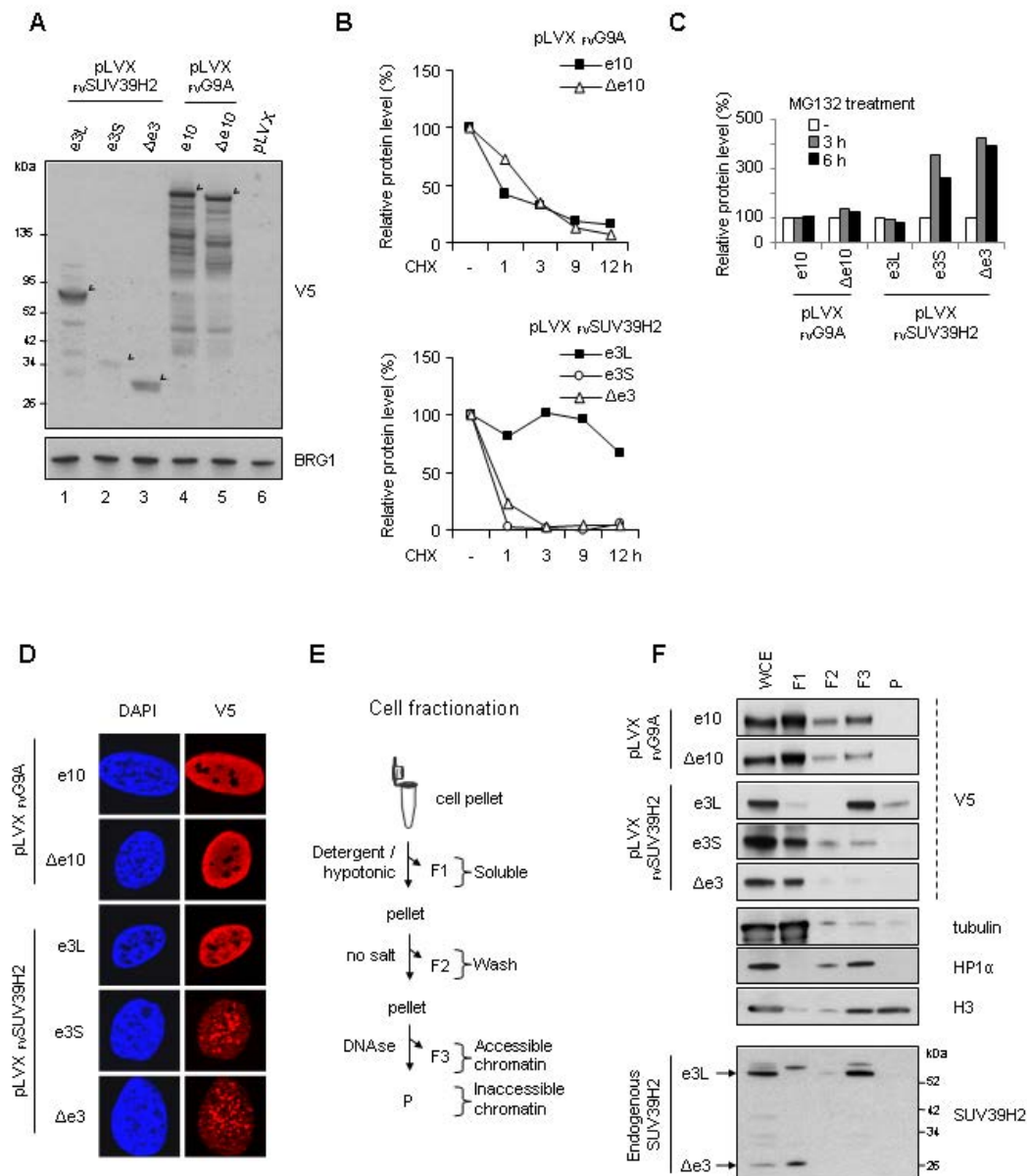
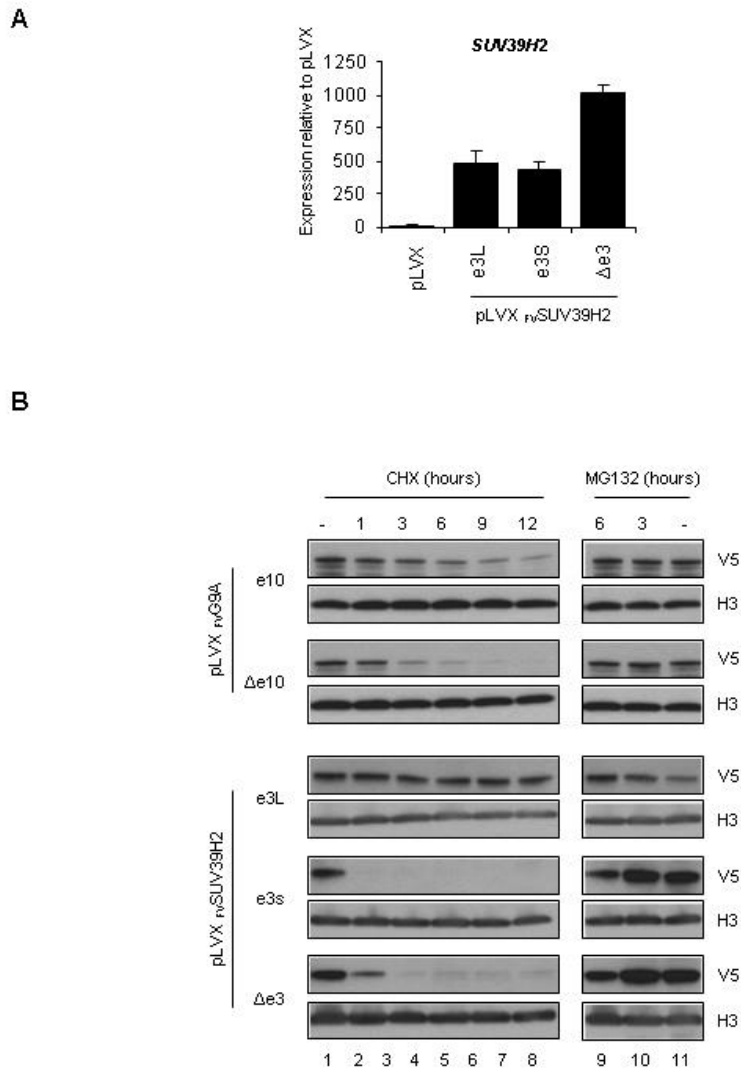


Figure 3

Figure 3. Transcripts resulting from alternatively spliced exon 3 of SUV39H2 encode proteins with stability and subnuclear localization differences. (A) Expression in HEK 293T cells of ectopic tagged G9A and SUV39H2 protein isoforms. Cells were transduced with the pLVX vector containing Flag-V5 tags and *SUV39H2* or *G9A* cDNA to express tagged forms of the proteins, *FvSUV39H2* or *FvG9A*, respectively. The ectopic proteins were detected by Western blot using the V5 antibody. (B) and (C) Analysis of the stability of the protein encoded by each G9A and SUV39H2 isoforms. The expression levels of ectopic *FvG9A* and *FvSUV39H2* protein isoforms was assessed by Western blot in HeLa cells untreated (-) or treated with cycloheximide (CHX) (B) or MG132 (C) for the indicated times in hours (h). The signal detected with V5 antibody was normalized to histone H3 levels. (D) Localization of ectopic *FvG9A* and *FvSUV39H2* protein isoforms expressed in HeLa cells. The V5 antibody (red) was used to detect the ectopic factors, while DNA was counterstained with DAPI (blue). (E) and (F) Cell fractionation of HeLa cells to analyze the distribution of ectopic *FvG9A* and *FvSUV39H2* and the endogenous SUV39H2 protein isoforms. (E) Scheme displaying the cell fractionation procedure. (F) The various protein isoforms were detected in each fraction and whole cell extracts (WCE) by Western blot using the V5 or SUV39H2 antibodies. The tubulin, HP1α and H3 proteins were also analysed as markers of the cell fractions.



Supplementary figure S3

Supplementary Figure S3. Transcripts resulting from alternative spliced exon 3 of SUV39H2 encode for proteins with stability and subnuclear localization differences. (A) *SUV39H2* transcripts expression in cells transduced with pLVX _{FVSUV39H2}_e3L, e3S, Δe3 or with the empty pLVX as negative control. (B) Analysis of the protein stability encoded by each G9A and SUV39H2 isoforms. The expression level of exogenous _{FVG9A} and _{FVSUV39H2} protein isoforms was assessed by Western blot in HeLa cells untreated (-) or treated with cycloheximide (CHX) (B) or MG132 (C) for the indicated times in hours (h).

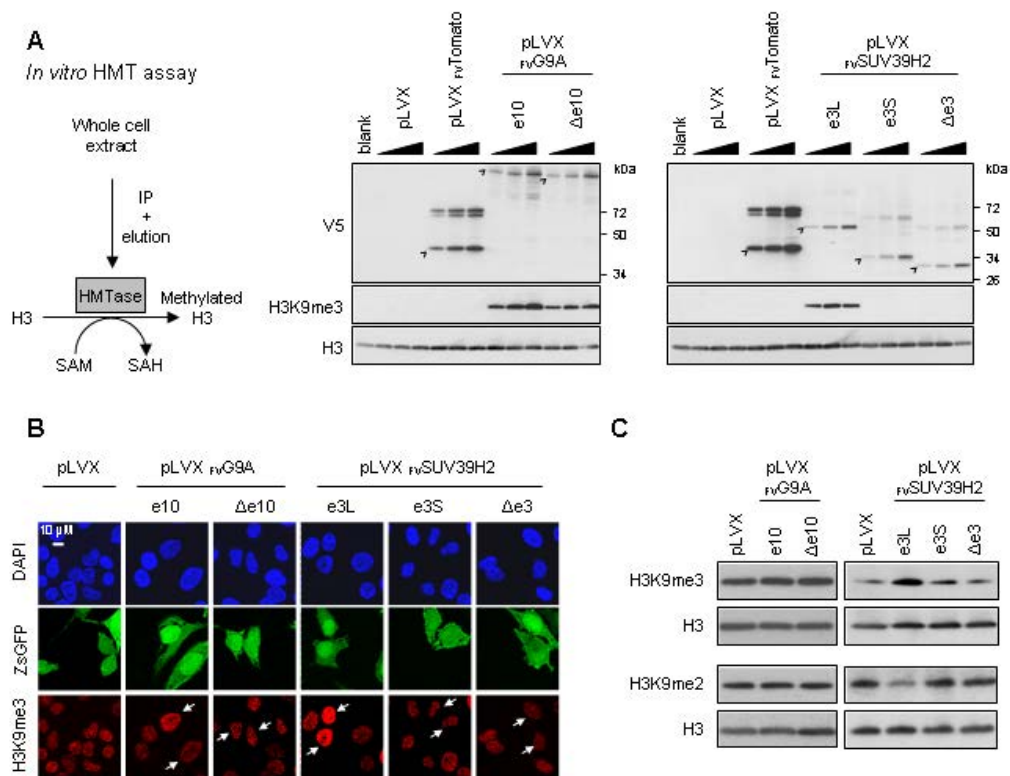
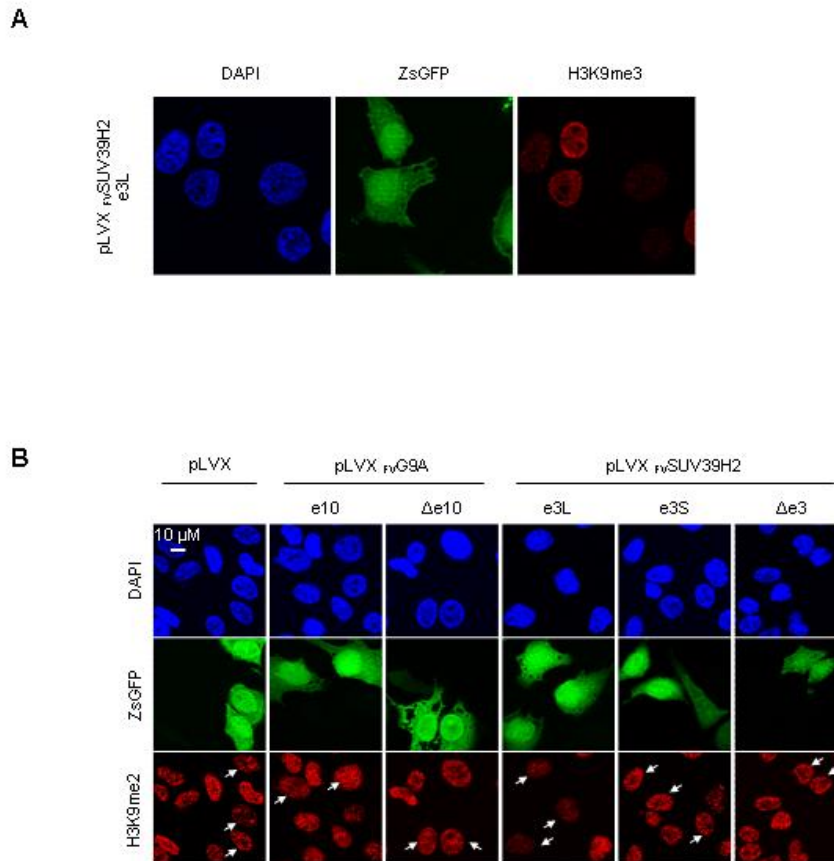


Figure 4

Figure 4. Inclusion of exon 3 in SUV39H2 transcripts is required to encode an active histone methyltransferase. (A) *In vitro* analysis of the methyltransferase activity associated to each G9A and SUV39H2 protein isoforms. Scheme of the *in vitro* histone methyltransferase (HMT) assay procedure (left panel). Various quantities of purified *FV*G9A and *FV*SUV39H2 enzymes (right panel) were incubated with the human histone H3.1 and the S-adenosylmethionine (SAM) producing methylated H3.1 and S-adenosylhomocysteine (SAH). The methylation reactions were analyzed by Western blot using H3K9me3 and H3 specific antibodies (right panel). Cells transduced with empty vector (pLVX) or cells transduced with a FLAG-V5 tagged version of tomato (*FV*Tomato) were used as controls. V5-tag detection was used to control for the amounts of HMTases used. (B) and (C) *In vivo* detection of the lysine 9 trimethylation of histone H3 (H3K9me3) in HeLa cells after expression of the different ectopic *FV*G9A and *FV*SUV39H2 protein isoforms or empty vector (pLVX) as control. (B) The transduced cells were identified with the bicistronic expression of the ZsGreen1 Fluorescence Protein (ZsGFP panel) and labelled with white arrows in H3K9me3 panel. H3K9me3 levels were estimated by immunostaining (red) and DNA was counterstained with DAPI (blue). (C) Analysis by Western blot of H3K9me3, H3K9me2 and H3 levels in total protein extracts of HeLa cells transduced to express *FV*G9A and *FV*SUV39H2 protein isoforms.



Supplementary figure S4

Supplementary Figure S4. Inclusion of exon 3 in SUV39H2 transcripts is required to encode for an active histone methyltransferase. *In vivo* detection of the lysine 9 trimethylation of histone H3 (H3K9me3) (non saturated picture compared to Figure 4B) (A) and lysine 9 dimethylation of histone H3 (H3K9me2) (B) in HeLa cells after expression of exogenous FV_{G9A} and FV_{SUV39H2} protein isoforms. As control, this experiment was also performed with cells transduced with empty vector (pLVX). The transduced cells were identified with the bicistronic expression of the ZsGreen1 Fluorescence Protein (ZsGFP panel) and labelled with white arrows in H3K9me2 panel. H3K9me2 levels were estimated by immunostaining (red) and DNA was counterstained with DAPI (blue).

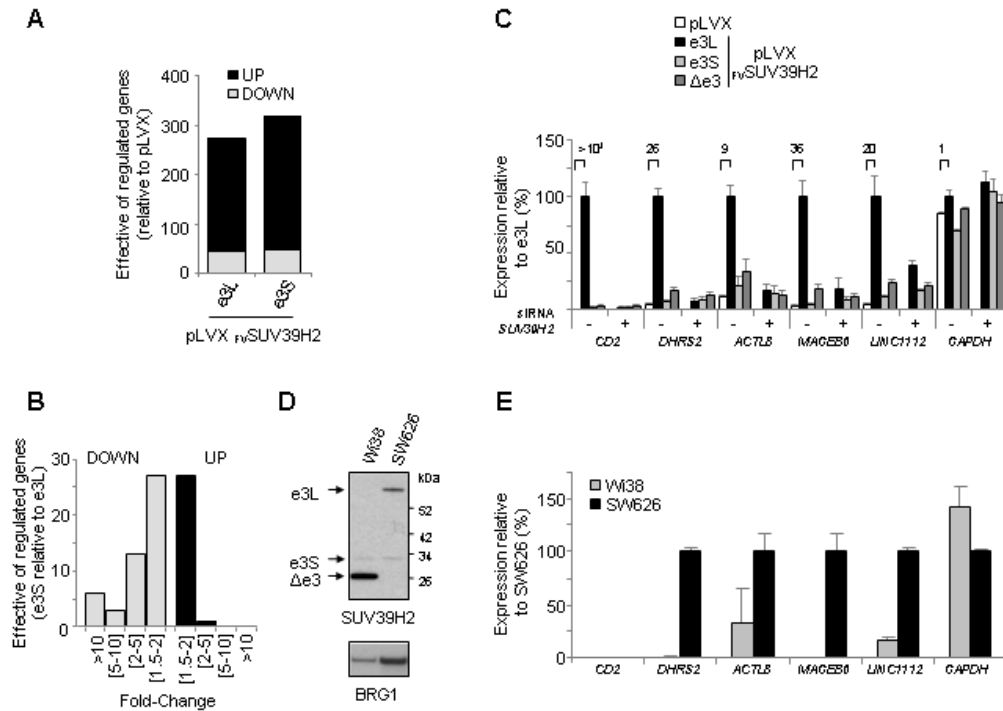
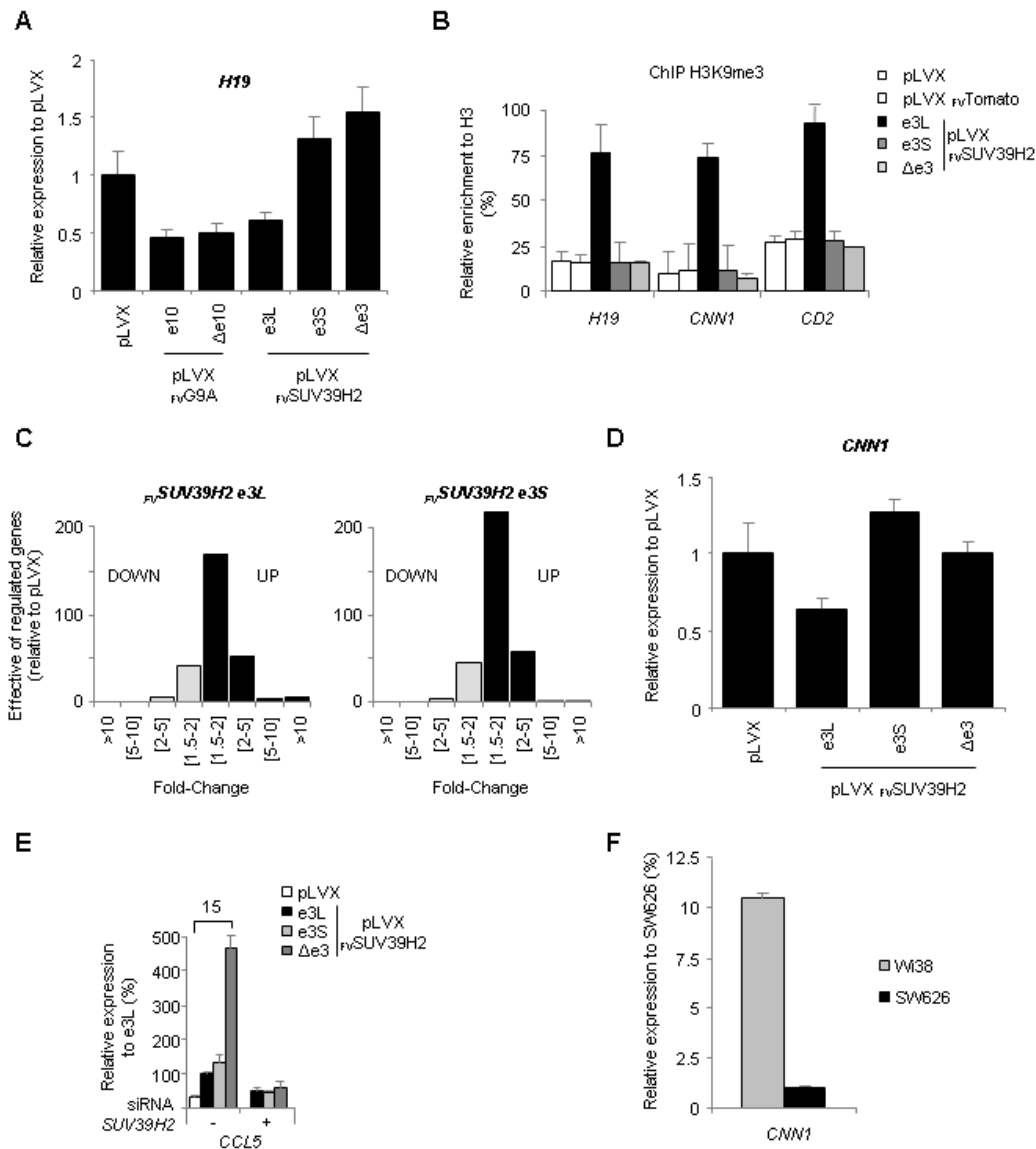


Figure 5

Figure 5. Transcriptional regulation of different target genes by SUV39H2 protein isoforms. (A) Graph showing the number of genes that are up- or down-regulated in HeLa cells after transduction with pLVX FVSUV39H2_e3L or pLVX FVSUV39H2_e3S. The mRNA expression levels were assessed by high throughput RNA sequencing and hits were retained whenever their expression fold-change was 1.5 higher than control (pLVX). (B) Graph displaying the number of genes differentially expressed in cells expressing ectopic FVSUV39H2_e3S compared to those expressing FVSUV39H2_e3L according to their fold-change. (C) Validation by RT-qPCR of relative gene expression for 5 hits differentially regulated by each SUV39H2 protein isoforms. The significance of this validation was assessed comparing results obtained from HeLa cells transfected (+) or not (-) by siRNA targeting all *SUV39H2* transcripts. Expression fold-change between cells transduced with pLVX FVSUV39H2_e3L and pLVX are indicated above the graph. (D) Western blot detection of endogenous SUV39H2 protein isoforms expressed in Wi38 and SW626 cell lines. (E) RT-qPCR analysis of relative gene expression in Wi38 and SW626 cells of the 5 hits validated in panel 5C. CD2 gene expression is not detectable in both cell lines.



Supplementary figure S5

Supplementary Figure S5. The protein isoforms of SUV39H2 regulate the transcription of different target genes. (A) RT-qPCR of relative gene expression for *H19* in HeLa cells transduced to express G9A and SUV39H2 isoforms. (B) H3K9me3 signal at *H19*, *CNN1* and *CD2* promoters detected by ChIP experiments in HeLa cells transduced to expressed SUV39H2 isoforms. (C) Graph showing the number of genes that are up- or down-regulated in HeLa cells after transduction with pLVX_{rvSUV39H2_e3L} or pLVX_{rvSUV39H2_e3S}. The mRNA expression levels were assessed by high throughput RNA sequencing and hits were retained whenever their expression fold-change was 1.5 higher than control (pLVX). (D) RT-qPCR of relative gene expression for *CNN1* in HeLa cells transduced to express SUV39H2 isoforms. (E) Validation by RT-qPCR of relative gene expression for *CCL5* differentially regulated by each SUV39H2 protein isoforms. The significance of this validation was assessed comparing results obtained from HeLa cells transfected (+) or not (-) by siRNA targeting all *SUV39H2* transcripts. Expression fold-change between cells transduced with pLVX_{rvSUV39H2_e3L} and pLVX are indicated above the graph. (F) RT-qPCR analysis of relative gene expression in Wi38 and SW626 cells of *CNN1*.

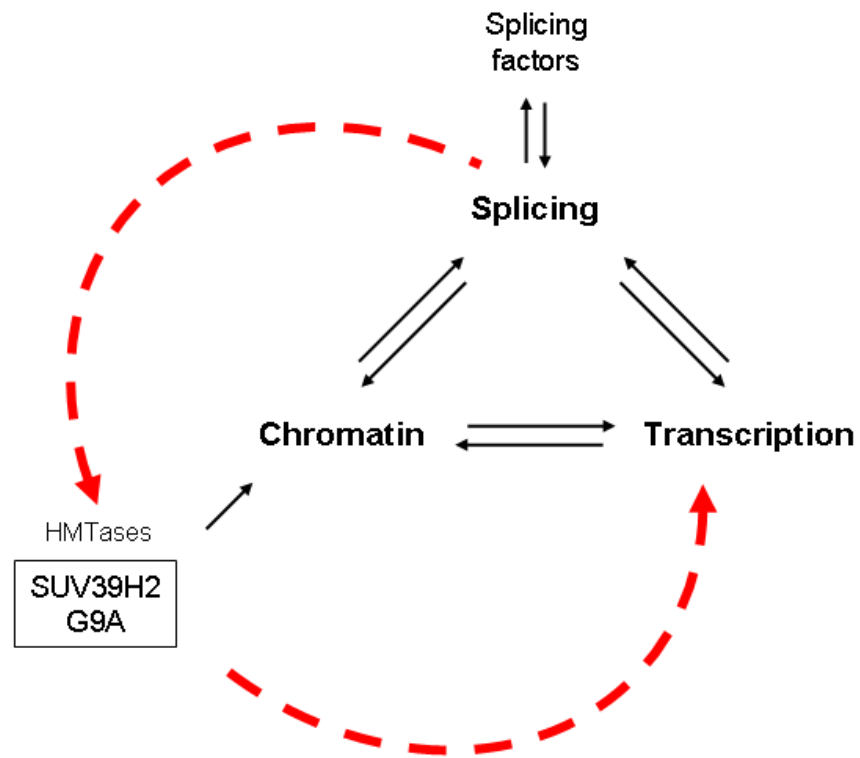
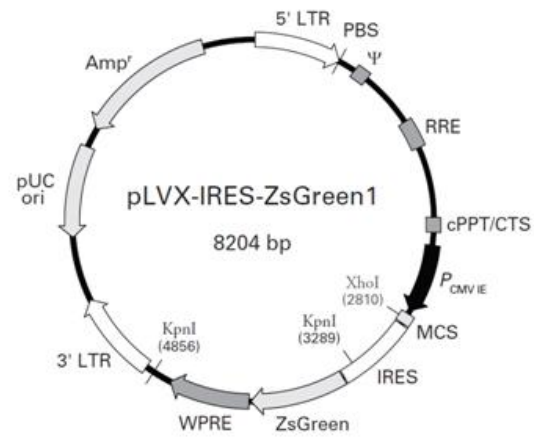


Figure 6

Figure 6. Coupling between splicing, chromatin and transcription.

A



Supplementary figure S6

G9a cDNA (ENST00000375537) cloned in pLVX lentiviral vector

GAATTCCTCGAGACTAGTATCTTCCACCATGGATTACAAGGATGACGACGATAAGGGTAAGCCTATCCCTAACCCCTCTCTCGGTCTCGATT
CTACGGGATCCCCGGGCTGCGAAGTTCAAGCTTACTCGAGATGGCGGCGGGCGGGAGCTGCAGCGGCGGCGGGCGGGAGGGGAGGC
CCCCGCTGAGATGGGGGCGTGTCTGGAGAAGGAAACCAGAGGAGCCACCGAGAGATTTCATGGCTCTTTGGGGGACACCCCTCGTAGTGAA
GAAACCCCTGCCAAGGCCACCCCGACTCCCTGGAGCCTGCTGGCCCCATCTCCAGCCTCTGTCACTGTCTGTTGGTGATGAGGGGGCTG
ACACCCCTGTAGGGGTACACCACTCATTGGGGATGAATCTGAGAATCTTGGGGAGATGGGGACCTCCGTGGGGGCGGATCCTGTGGGCCA
TGCCACAAAGTCATTCCTCTTCCCCCAGCAAGGGGGTTCTGTCTAGCCGGGCCAAGATGTCAATGACAGGGGCGGAAAAATCACCTCCA
TCTGTCCAGATTGGCTATGAGGCTACTGAGTATGCCAGGAGCCAGGGAGCTGCAGCAGCAGGGTCTGAACCCCTCCAGCCACCACGAGCC
CAGAGGGACAGCCCAAGGTCCACCGAGCCCGCAAAACCATGTCCAAACCAGGAAATGGACAGCCCGCTCCCTGAGAAGCGGGCCCTGAAAT
ACAGCATTTCGCATGAGTGTGATGTCCACTCACTGGGAAAGGTGACCTCAGATCTGGCCAAAAGGAGGAAGCTGAACCTCAGGAGGTGGCCTG
TGGAGAGTTAGGTTCTGCCCCGGCTTCAGGAGAAGTGACCTGACGAAAGGGGACCCCGGGTCCCTGGAGGAGTGGGAGACGGTGGTGGGTG
ATGACTTCAGTCTCTACTATGATTCTACTCTGTGGATGAGCGGTGGACTCCGACAGCAAGTCTGAAGTTGAAGCTCTAACTGAACAATAAG
TGAAGAGGAGGAGGAGGAAGAGGAGGAAGAAGAGAGGAAGAGGAGGAGGAAGAGGAAGAAGAGGAAGATGAGGAGTCAAGGGAATCAG
TCAGATAGGAGTGGTTCAGATGGCGGGCGCAAGGCCAAGAAGAAATGGCGAAAAGACAGCCCATGGGTGAAGCCGTCTCGGAAACGGCGCAAGC
GGGAGCTCCCGCGGGCCAAGGACCCACGAGGAGTGAATGGTGTGGGTCTCTCAGGCCCCAGTGAATGAGTACATGGAGGTCTCTGGGGTCCCTGGA
GCTGCCACGAGGGGACCCCTCTCCCCAACCCACGCTGGGGTGTCCATGACACATCTTCGCTGGAGACAGAGCGAGGGTTTGGAGGTTGCC
CTGTGACGCTGCCGCATGGAGGACCCCAAGATTGACCCCATCAGCGAGAGGGCGGGGACAAAGTGCATGGCCACTGAGAGTGTGGACGGAGAGC
TGTCAGGCTGCAATGCCGCCATCCTCAAGCGGGAGACCATGAGGCCATCCAGCCGTGTGGCCCTGATGGTGTCTGTGAGACCCACCGCGCCCG
CATGGTCAAAACACCATCTGCTCCCGGGTCTGCGCTACTTCTGACAGCTGCGGACCCCTTCTGAGAGTGCACCCCTGATCCGTGAGGAGTGC
TTCCACAAGGCCCTGTGTCTCAGCTGAATGGGATGGTCTTCTGTCCCCACTGTGGGGAGGATGCTTCTGAAGCTCAAGAGGTGACCATCCCC
GGGGTGACGGGGTGACCCACCGCGCGGCACTGCAGCTCCTGCACCCCAACCCCTGTCCAGGATGTCCCGGGAGAGCAGACACTTCTCAGCC
CAGTGCCCGGATGCGAGGGCATGGGGAACCCCGCGCGCCCTGCGATCCCCCTGGCTGACACCATGACAGCTCAGGGCCCTCCCTGACCCCTG
CCCAATGGGGTGTGCTTCCAGCCGTGGGCTGCCACTGGGCGGAGCCCGGGAGGCCCTGGAAAAGGCCCTGGTATCCAGGAGTCAAGAGAGC
GGAAGAAGCTCCGTTTCCACCCTCGGCAGTTGTACCTGTCCGTGAAGCAGGGCGAGCTGCAGAAGGTGATCCTGATGCTGTTGGACAACCTGGA
CCCCAATCTCCAGAGCGACAGCAGAGCAAGCGCACGCCCTGATGCAAGCGGCCAGAGGGCTCCGTGGAGATCTGCCATGTGCTGTGTCAG
GCTGGAGCCAACATAATGCAGTGGACAACAGCAGCGGACGCCACTGATGGAGGCGGTGGTGAACAACACCTGGAGGTAGCCCGTTACATGG
TGCAGCGTGGTGGCTGTGCTATAGCAAGGAGGAGGACGTTTCCACCTGCCCTCCACACGACGCAAAATCGGGAACTTGGAGATGGTCAGCCT
GCTGCTGAGCACAGGACAGGTGGACGTCAACGCCAGGACAGTGGGGGTGGACGCCCATCATCTGGGCTGCAGAGCACAAGCACATCGAGGTG
ATCCGCATGCTACTGACGCGGGGCGCCGACGTACCCTCACTGACAACAGGAGAAACATCTGCCTGCCTGAGGCTCCCTTACCGGCGAGCGCCG
CCATCGCCGAAGTCTTCTGAATGCGCGCTGTGACCTCCATGCTGTCAACTACCATGGGGACACCCCCCTGCACATCGCAGCTCGGGAGAGCTA
CCATGACTGCGTGTCTTATTCCTGTACGCTGGGGCCCAACCTCAGCTGCGGAAACAAAGAGGGGACACAGCATGGGACCTGACTCCCGGACGC
TCCGACGTGTGGTTTGCCTTCAACTCAACCGCAAGCTCCGACTTGGGGTGGGAAATCGGGCCATCCGCACAGAGAAGATCATCTGCCGGGACG
TGCTCGGGGCTATGAGAACGTGCCATTCCTGTGTCAACCGTGTGGATGGGGAGCCCTGCCCTGAGGATTACAAGTACATCTCAGAGAACTG
CGAGACGTCCACCATGAACATCGATCGCAACATCACCCACCTGCAGCACTGCAGTGTGTGGACGACTGCTCTAGCTCCAACCTGCCTGTGCGGC
CAGCTCAGTACATGCTGTGATGACAAGGATGGGCGATTGCTCCAGGAATTTAACAAGATTGAGCCTCCGCTGATTTTCAGTGTAAACGAGG
CGTGCTCATGCTGGAGAACTGCAAGAACCAGGCTCGTACAGAGTGGCATCAAGGTGCGGCTACAGCTCTACCGAACAGCAAGATGGGCTGGGG
GGTCCGCGCCCTGCAGACCATCCACAGGGGACCTTCACTGTGCGAGTATGTGCGGGAGCTGATCTCTGATGCTGAGGCTGATGTGAGAGAGGAT
GATTCTTACCTCTTCGACTTAGACAACAAGGATGGAGAGGTGTACTGCATAGATGCCCCGTACTATGGCAACATCAGCCGCTTCAACACACC
TGTGTGACGCTCCACATATTCCTGTCGCTCCGGTCTTCATGCTGCACCAAGACCTGCGGATTTCCACGCACTCGCTTCTCAGTTCCCGAGACATCCG
GACTGGGGAGGAGCTAGGGTTTACTATGGCGACCGCTTCTGGGACATCAAAAGCAAATATTTACCTGCCAATGTGGCTCTGAGAAGTGCAAG
CACTCAGCCGAAGCCATTCGCTTGGAGCAGAGCCGTCTGGCCCGCTTGACCCACACCCCTGAGCTGCTGCCCGAGCTCGGCTCCCTGCCCCCTG
TCAACACAATGATCGACCTCGAGG

MCS ; Kozac ; FLAG ; V5 ; G9a cDNA ;
Alternative exon 10 absent in pLVX_{FV}G9a_Δe10

SUV39H2 cDNA (ENST00000354919) cloned in pLVX lentiviral vector

GAATTCCTCGAGTCCACCATGGATTACAAGGATGACGACGATAAGGGTAAGCCTATCCCTAACCCCTCTCTCGGTCTCGATTCTACGGGATCC
TCTACAAGATGGCGGCGGTGCGGGCCGAGGCGCGAGGAGCTTGGTGTGCTTGCCTAGTTTCACTTGATACTCTTCAGGAATTATGTAGAAA
AGAAAAGCTCACATGTAATCGATTGGAATCACCAAAAGGAATCTAAACAATTATGAGGTGGAATACTTGTGTGACTACAAGGTAGTAAAGGAT
ATGGAATATTATCTTGTAAAATGAAAGGATGGCCAGATTCTACAAATACTTGGGAACCTTTGCAAAATCTGAAGTGCCCGTTACTGCTTCAGC
AATTCTCTAATGACAAGCATAATTATTTATCTCAGTAAAGAAAGGCAAGCAATAACTCCAAAAGACAATAACAAAACCTTTGAAACCTGCCAT
TGCTGAGTACATTTGTGAAGAAGGCTAAACAAAGGATAGCTTCGACAGAGATGGCAAGATGAACCTCAACAGAAAGAAAGAAATCATAAAGGAATGATA
TTTGTGTAATACTGTTGATTAGAGGGCCACCTTCAGACTTCTATTACATTAACGAATACAAACCAGCTCCTGGAATCAGCTTAGTCAATG
AAGCTACCTTTGGTTGTTTCATGCACAGATTGCTTCTTTCAAAAATGTTTGTCTGCTGAAGCTGGAGTTCTTTTGGCTTATAATAAAAAACCAACA
AATTAATAATCCCACTGGTACTCCCATCTATGAATGCAACTCAAGGTGTGAGTGTGGTCTGATTGTGCCAATAGGATTGTACAAAAAGGCACAC
CAGTATTTCGCTTTGCATCTTTTCCGAAGTACGCAATGGACGTGGCTGGGGGTGAAAGACCCCTTGTGAAGATTAAAGAAATGAGTTTGTCTGATGAAT
ATGTTGGAGAGTAAATCACAAGTGAAGAAGCTGAAAGACGAGGACAGTTCTATGACAACAAGGAATCAGTATCTCTTTGATCTGGACTATGA
GTCTGATGAATTCACAGTGGATGCGGCTCGATACGGCAATGTGTCTCATTTTGTGAATCACAGCTGTGACCCAAATCTTCAGGTGTTCAATGTT
TTCATTGATAACCTCGATACTCGTCTTCCCCGAATAGCATTGTTTCCACAAGAACATAAATGCTGGAGAAGAGCTGACTTTTGATTATCAAA
TGAAAGGTTCTGGAGATATATCTTCAGATTCTATTGACCACAGCCAGCCAAAAGAGGGTGCAGAACAGTATGTAATGTGGAGCTGTGACTTG
CAGAGGTTACCTCAACGACGCGCGGATCG

MCS ; Kozac ; FLAG ; V5 ; SUV39H2 cDNA ;
Alternative exon 3 part absent in pLVX_{FV}SUV39H2_Δe3
Alternative exon 3 part absent in pLVX_{FV}SUV39H2_e3S and pLVX_{FV}SUV39H2_Δe3

Figure S6. Lentiviral vectors map and sequences. (A) pLVX-IRES-ZsGreen1 Vector Map (Clontech, #632187). **(B)** Sequences of G9a and SUV39H2 cDNAs cloned in pLVX vector.

FAST DB STABLE ID	Gene Symbol	Regulation	Fold- Change	P-Value	Intensity G9a+	Intensity G9a-
GSHG0036106	MIRLET7I	up	2,92	8,29E-03	4,46	13,05
GSHG0031492	INE1	up	1,97	2,16E-02	13,94	27,40
GSHG0018727	NNAT	up	1,93	7,28E-03	15,45	29,85
GSHG0014590	CNN1	up	1,90	5,90E-07	51,95	98,69
GSHG0023789	--	up	1,84	2,59E-02	15,05	27,67
GSHG0013280	CCL5	up	1,81	4,37E-04	38,14	69,16
GSHG0021365	--	up	1,69	4,56E-02	17,08	28,85
GSHG0025159	C5orf46	up	1,68	3,37E-03	24,05	40,40
GSHG0010003	ISLR	up	1,67	9,04E-05	49,41	82,73
GSHG0020267	LRP5L	up	1,67	2,41E-02	23,44	39,21
GSHG0011038	--	up	1,65	1,32E-02	22,68	37,45
GSHG0025255	KCNMB1	up	1,63	9,16E-04	44,50	72,50
GSHG0012400	CCL11	up	1,62	5,85E-03	34,48	55,92
GSHG0013210	VTN	up	1,60	6,58E-05	97,40	155,98
GSHG0024540	--	up	1,60	8,22E-03	33,00	52,81
GSHG0013393	KRT15	up	1,59	5,71E-03	41,66	66,15
GSHG0000343	--	up	1,56	4,89E-02	23,12	36,16
GSHG0029917	LY6D	up	1,55	2,54E-02	42,89	66,49
GSHG0022884	GUSBP5	up	1,55	3,42E-02	28,48	44,12
GSHG0033447	--	up	1,55	1,76E-02	39,96	61,75
GSHG0026126	--	up	1,54	2,22E-02	26,77	41,32
GSHG0029908	TSNARE1	up	1,54	1,97E-02	34,62	53,32
GSHG0019332	HELZ2	up	1,54	1,41E-02	38,04	58,50
GSHG0004938	RARRES3	up	1,53	1,13E-03	71,37	109,12
GSHG0024271	EGR1	up	1,51	5,61E-07	598,26	904,81
GSHG0015033	IGFL1	up	1,51	4,01E-02	37,65	56,93
GSHG0039495	CBX3P2	up	1,51	2,74E-02	24,40	36,88
GSHG0021739	ZMYND10	up	1,51	4,21E-02	34,14	51,44
GSHG0000845	CD2	down	Inf	2,50E-10	23,64	0,00
GSHG0025320	AACSP1	down	43,68	1,91E-09	25,25	0,58
GSHG0031438	MAGEB6	down	21,71	4,16E-35	84,51	3,89
GSHG0007126	--	down	13,53	7,74E-10	19,05	1,41
GSHG0000193	ACTL8	down	12,94	1,63E-24	118,33	9,14
GSHG0008734	DHRS2	down	11,23	2,17E-21	71,41	6,36
GSHG0029029	--	down	7,87	8,62E-25	104,42	13,27
GSHG0005810	TCN1	down	7,08	1,14E-24	104,62	14,77
GSHG0024786	--	down	5,24	4,49E-03	10,15	1,94
GSHG0023231	COX7B2	down	4,83	2,19E-04	14,78	3,06
GSHG0032096	PAGE1	down	3,19	1,97E-05	34,89	10,93
GSHG0003216	SUV39H2	down	3,02	2,22E-103	93338,64	30907,29
GSHG0029002	C8orf22	down	2,75	2,28E-03	30,03	10,91
GSHG0040123	LINC00665	down	2,55	1,41E-10	111,99	43,93
GSHG0035368	SNORD15B	down	2,50	2,70E-02	12,40	4,95
GSHG0016158	ZNF772	down	2,45	1,75E-02	15,31	6,26
GSHG0031893	CSAG1	down	2,35	2,54E-04	41,69	17,74
GSHG0027619	--	down	2,17	3,02E-04	48,73	22,45
GSHG0021423	TP63	down	2,17	3,09E-04	40,51	18,69
GSHG0031252	--	down	2,15	2,17E-02	19,12	8,87
GSHG0031214	FIBCD1	down	2,14	2,01E-06	100,95	47,18
GSHG0023034	--	down	2,01	2,64E-03	41,60	20,71

GSHG0022405	FGFR3	down	1,98	4,77E-02	36,96	18,65
GSHG0014861	--	down	1,90	2,94E-03	39,80	20,94
GSHG0007412	VDR	down	1,90	2,53E-07	119,68	63,02
GSHG0020652	RARB	down	1,86	4,09E-02	19,90	10,72
GSHG0010729	--	down	1,83	1,83E-09	169,03	92,38
GSHG0022297	ETV5	down	1,75	4,64E-05	99,05	56,61
GSHG0003375	DKK1	down	1,74	1,52E-04	95,93	55,06
GSHG0023928	FST	down	1,73	8,03E-04	62,10	35,95
GSHG0013836	NOTUM	down	1,72	3,87E-03	79,63	46,34
GSHG0026305	ELOVL2	down	1,71	2,60E-05	94,76	55,30
GSHG0008914	HSPA2	down	1,68	3,64E-09	256,36	152,85
GSHG0015899	ZNF229	down	1,67	5,32E-09	224,23	133,89
GSHG0031668	FAM133A	down	1,66	4,79E-07	347,20	208,88
GSHG0028479	CCDC71L	down	1,63	9,91E-08	223,80	137,02
GSHG0001444	KCNK2	down	1,61	9,54E-05	144,44	89,91
GSHG0045802	CAMK2N1	down	1,58	9,31E-06	183,56	115,92
GSHG0022555	UCHL1	down	1,57	6,73E-09	342,02	217,40
GSHG0009182	TMEM121	down	1,57	1,16E-02	53,53	34,03
GSHG0026645	NFKBIE	down	1,57	7,24E-03	56,41	36,03
GSHG0005864	RCOR2	down	1,56	2,58E-02	45,55	29,11
GSHG0009996	NEO1	down	1,55	4,38E-03	129,81	83,58
GSHG0017150	MAP2	down	1,55	1,51E-03	113,82	73,52
GSHG0027948	ADAP1 // COX19	down	1,54	3,71E-02	36,81	23,83
GSHG0029420	LONRF1	down	1,54	1,32E-02	37,99	24,65
GSHG0031133	--	down	1,51	2,11E-04	117,12	77,33
GSHG0012534	TMEM106A	down	1,51	4,74E-02	35,63	23,57
GSHG0017713	TGFA	down	1,51	4,31E-05	200,10	132,71

Table S1. Genes differentially expressed according to enforced expression of SUV39H3_E3S versus SUV39H2_E3L

Gene Name	Forward primer sequence	Reverse Primer sequence
human G9a	GCAAGGCCAAGAAGAAATGGCGAA AGTGATGATGTCCACTCACTGGGA	TCTCGCTGATGCGGTCAATCTT AGAGACTGAAGTCATCACCCACCA
mouse G9a	GCGCAAGGCCAAGAAGAAATG	CAGTGAAGACGTGTCATTGGAGAC
human SUV39H2	GTGTGTGCCTTGCCTAGTTTCACT ATGAATTCACAGTGGATGCGGCTC	CTGTCTCGTCTTTTCAGCTTCTTCAC TGGCTGGGCTGTGGTCAATAGAAT
mouse SUV39H2	GCGCAAGGCCAAGAAGAAATG	CAGTGAAGACGTGTCATTGGAGAC
chicken SUV39H2	GTATGTGCCATGTCTGGCTTCAC	CGTCTCTCTGCTTCCTCACTTGT
zebrafish SUV39H1b	CACAGATGGGAAGAGGCAAACA	CGTCCACATAATCCAGGTCAAACA
human RLP0	AGGTGTTTCGACAATGGCAGCAT	TGCAGACAGACACTGGCAACAT
human GAPDH	TCCCATCACCATCTTCCAGG	CATCGCCCCACTTGATTTTG
human LINC1112	AGAAATATACAGGCATACTTGTGG	TGCACTGTATTTTTCCACATGTTG
human CD2	CAGGGAACAAAGTCAGCAAGGAA CTGTGTGGTTGAATTTGGCTTCTTG	CAGTGCCACAAAGACCATCAAGAG GCCCTTCTGATGTGCTCCTTT
human DHRS2	CCAGTGAGCAGATCTGGGACAA	TTCTCCATGTAGGGCAGCAACT
human ACTL8	ATCATCACGGAGACACCCTTGA GCTGCTTTGCGTCTGGTTTCT	CATAGCCAGAATCAACCACCACTC ATGTGGAGCGGAGAGGCAATAA
human MAGEB6	GTCCCAGATCTTCTCAGCTGTCTT CAACTCTGGACTCTCAGGGAATGA	CTCACAGGTACGGAGCTTACTCTT GGACCCTTGGACTTCTCTATCTGTT
human CCL5	CGTGCCACATCAAGGAGTATTT CTCAGTGACTTCTGATCCTGTCCTAAC	CAAGGACTCTCCATCCTAGCTCAT TCTCCCTCACTGCTCTCTCATCTA
human CCN1	ACATGTGAGGAGGGAAGAGTGT CCTGCTTCAAGACTCCAGAGAGATA	TCGTTGAAGTGAGCAGAGGA GTGATGTCAGGCCCTTGGTATTG
human H19	AGTGGACTTGGTGACGCTGTAT	TGGATGCTGTACTGTCTGCCAA

Table S2. Primers sequences

Name	name strand	siRNA location	siRNA 5'→3'	overhanged
Control	sense antisense	non targetting	UAGCAAUGACGAAUGCGUA UACGCAUUCGUCAUUGCUA	dTdT dTdT
G9a_1	sense antisense	ex11-ex12	GGAGAGCUGUCAGGCUGCA UGCAGCCUGACAGCUCUCC	dTdT dTdT
G9a_2	sense antisense	ex24-ex25	GCAUCAAGGUGCGGCUACA UGUAGCCGCACCUUGAUGC	dTdT dTdT
SUV39H2_1	sense antisense	ex4-ex5	CAGCUGUGACCCAAAUCUU AAGAUUUGGUCACAGCUG	dTdT dTdT
SUV39H2_2	sense antisense	ex5-ex6	GAAAGGUUCUGGAGAUUAU UAUAUCUCCAGAACCUUUC	dTdT dTdT

Table S2 bis. siRNA location and sequences

Sample	Raw reads	Average base quality over 30	Sequence length	%GC	Total mapped reads
pLVX_1	57,683,411	100%	50	51	(93.5%) 53,336,056
pLVX_2	55,270,068	100%	50	50	(96.1%) 52,723,331
pLVX_3	42,509,141	100%	50	50	(96.2%) 40,592,080
G9a_e10_1	39,316,485	100%	50	51	(93.0%) 36,084,509
G9a_e10_2	56,272,391	100%	50	50	(96.4%) 53,851,594
G9a_e10_3	29,995,057	100%	50	50	(96.6%) 28,768,149
G9a_Δe10_1	30,951,069	100%	50	51	(96.8%) 29,729,406
G9a_Δe10_2	44,665,122	100%	50	50	(96.7%) 42,872,820
G9a_Δe10_3	47,980,571	100%	50	50	(96.9%) 46,160,588
SUV39H2_e3L_1	35,387,857	100%	50	50	(96.2%) 33,809,030
SUV39H2_e3L_2	36,529,662	100%	50	50	(96.0%) 34,815,356
SUV39H2_e3L_3	65,932,495	100%	50	50	(96.0%) 62,871,033
SUV39H2_e3S_1	26,309,193	100%	50	50	(96.2%) 25,122,569
SUV39H2_e3S_2	49,701,123	100%	50	51	(96.2%) 47,465,352
SUV39H2_e3S_3	53,407,232	100%	50	50	(96.2%) 51,009,422

Table S3. Illumina sequencing run statistics

DISCUSSION

Chromatin modifiers (alter. name)	Functions of the main product	Isoforms	Functions of the alternative spliced isoforms	References
CARM1 (PRMT4)	Arg methylase H3R8	4 isoforms in human, mouse & rat 2 isoforms in <i>X. laevis</i>	CARM1-v3 regulates alternative splicing of CD44 and E1a Exon15 encodes an automethylated domain of the protein	(Matsuda et al., 2007; Ohkura et al., 2005; Wang et al., 2013)
DNMT3B	DNA methylation	40 isoforms DNMT3B4 B5 lacking ex21+22 DNMT3B3-like splice variant lacking exon 5	Exons21+22 encoding catalytic domain Exon5 increases DNA binding or influences the 3D structure of the adjacent PWWP domain	(Robertson et al., 1999; Gopalakrishnan et al., 2009; Van Emburgh and Robertson, 2011)
SIRT1	NAD-dependent deacetylase H4-K16 and H3-K9 deacetylation	SIRT1-ΔExon8 SIRT1-Δ2/9	SIRT1- Δ Exon8, displays significant differences in stress-sensitivity, RNA/protein stability, protein-protein interactions and deacetylase activity -Δ2/9 lacks catalytic domain	(Pruitt et al., 2006; Lynch et al., 2010; Shah et al., 2012)
PKM1/2	Pyruvate kinase (Warburg effect)	2 isoforms : PKM1 and PKM2	PKM2 phosphorylates H3T11p	(Yang et al., 2012)
LSD1	H3K4 demethylase	4 isoforms	4 aa exon E8a brings the phosphorylated Thr important for protein conformation and activity	(Zibetti et al., 2010; Toffolo et al., 2013)
EZH2	H3K27me3 HMTase	2 isoforms EZH2α EZH2β brings Exon4 skipping and Exon83'SS	EZH2β represses a predominantly unique subset of gene targets from EZH2α	(Grzenda et al., 2013)

Table 1. List of known chromatin factors regulated by alternative splicing.

Here, we report that the exon 10 of *G9A* and exon 3 of *SUV39H2* are two alternative spliced exons. Their splicing regulation is conserved through different species, tissues and cell lines supporting a functional link in the modulation of both enzymes. The alternative spliced variants encode for proteins and regulate the stabilities, the sub-nuclear localizations and HMTase activity of SUV39H2, unlike G9A. Furthermore, using high-throughput RNA sequencing, we found that the splice variant level of SUV39H2 involved in the expression of almost 80 genes. Our data suggest also that alternative exon 10 may control of a non-H3K9me

function of G9A, while exon 3 of SUV39H2 plays an essential role for its functional HMTase activity.

Study of chromatin factors need to take into account the alternative protein isoforms

Overall, our data argue that the functional study of chromatin factors should not be restricted to a single transcript. To our knowledge, only few studies about chromatin factors have investigated the functional impact of alternative splicing (**Table 1**). Here, we demonstrate for SUV39H2 an absence of relationship between its enzymatic activity and total mRNA level. Indeed, we found that its HTMase activity is dependent of exon 3, which is broadly subjected to a regulation by alternative splicing in 21 human tissues (**Figure 1A**). Supporting this view, we found a large part of the *G9A* and *SUV39H2* transcripts that do not include the variant exons. Therefore, the studies, which disregard alternative transcripts, lead to partial information about gene function.

Impact of alternative splicing on SUV39H2 expression and function

Interestingly, the expression of *SUV39H2* was detected in all tissues, while it was described initially as specific of testis (Carroll et al., 2000). In agreement with ubiquitous expression of *SUV39H2*, the *Suv39h1* ^{-/-} and *Suv39h2* ^{-/-} mice exhibit a stronger phenotype than the *Suv39h1* ^{-/-} mice, suggesting a larger redundant activity than only in testis tissue. In addition, the discrepancy between both studies could reflect differences in the procedure used to evaluate gene expression. In fact, the initial study of O'Carroll *et al.* have assessed *SUV39H2* expression by Northern Blot using a probe targeting exon 3, which led them to miss a large part of transcripts. Thus, these differences underline the importance to take into account alternative transcripts.

Moreover, we found that the level of alternative SUV39H2 transcripts and proteins are not linear, unlike G9A isoforms. For instance, *SUV39H2_Δe3* transcript is less abundant in IMR90 and WI38 cells, whereas it is the most expressed protein **isoform** (**Supplementary Figure S1C and S2A**). This is a proof of concept that beyond introducing diversity in protein sequence, alternative splicing is also a strong pathway to regulate the level of expression. Finally, the alternative versions of SUV39H2 encoded by *SUV39H2_Δe3* enhance strongly the expression of CCL5 compare to *SUV39H2_e3S* and *SUV39H2_e3L* (**Supplementary Figure S5E**). This may reflect that each alternative versions of SUV39H2 could carry its own activity on transcription regulation.

Inclusion of exon 10 in G9A transcript is not required for H3K9 methyltransferases activity.

Here, we report that the splicing regulation of *G9A* exon 10 is conserved between human and mouse. This is not relevant of a specific modulation in the properties of G9A assayed in our study (H3K9 methylation activity, protein stability, sub-nuclear localization). The absence of differences may reflect various explanations. The limit of approaches used to probe G9A activity; in particular they could not be sensitive enough to reveal the function of G9A on specific loci. An other possibility, exon 10 could encode for a protein motif controlled by signalling pathways, and *in fine* regulates G9A activity (Buljan et al., 2013; Ellis et al., 2012). Exon 10 is predicted to encode for an unstructured protein motif enriched in serine and threonine, which depict potential sites for phosphorylation. Finally, the activity of G9A was only tested for H3K9, because it is the most described substrate. Therefore, the absence of differences may be only true for H3K9, and impact on H3K27, H3K56 will have to be tested in future (Tachibana et al., 2001; Yu et al., 2012). Exon 10 could also encode for a protein domain that lead methylation on non-histone proteins (Rathert et al., 2008). Furthermore, we cannot exclude that exon 10 would be important for a still unknown activity of G9A.

SUV39H2 and G9A do not use the same pathway to methylate H3K9.

In vitro, G9A exhibits a strong H3K9 methyltransferases activity, while in HeLa cells, its overexpression doesn't impact the amount of H3K9me3, nor H3K9me2 (**Figure 4 and Supplementary Figure S4**). Such discrepancy may reflect the requirement of cofactors *in vivo*. A candidate could be GLP, an other HMTase interacting physically and functionally *in vivo* (Mozzetta et al., 2013; Tachibana et al., 2005). Corroborant this possibility, the gene silencing activity of G9A was shown to belong into complexes with SETDB1, SUV39H1 and the Polycomb complex (Fritsch et al., 2010; Mozzetta et al., 2013). In contrast, the activity of SUV39H2 was detectable *in vitro* and *in vivo*, supporting an absence of limiting cofactors *in vivo*. Interestingly, the differences observed between G9A and SUV39H2 suggest also that these enzymes do not modify H3K9 through redundant pathway *in vivo*.

High levels of SUV39H2 enhances transcription of target genes

The presence of H3K9me3 at the promoter is abundantly described in literature as a mark of transcriptional repression, thus it was unexpected to find an activation of transcription in HeLa cells after increase in level of active SUV39H2. We cannot exclude indirect effects led

by our technical approaches. Indeed, exogenous SUV39H2 could compete with the transcriptional repressors interacting with H3K9, as it was shown for SUV39H1 (Aagaard et al., 1999; Melcher et al., 2000; Sewalt et al., 2002). On the other hand, H3K9me3 mark is also found in the body of genes actively transcribed (Vakoc et al., 2005, 2006; Wiencke et al., 2007) and H3K9 methylation is bound by epigenetic activator, such as KDM4C (Gregory and Cheung, 2014). Therefore, we cannot rule out that our results reflect some physiological pathway. Supporting this view, the up-regulation of 65 genes was strictly dependent of the methyltransferase activity of SUV39H2 (**Figure 5B**). In addition, only CD2 showed an increase of H3K9me3 in its promoter, suggesting that other hits may be up-regulated through an effect of methylation in gene body (**Figure 5C and Supplementary Figure S5B**).

Splicing regulation of chromatin factors extends crosstalk between machineries

Since 30 years, coupling mechanisms have been extensively reported for interconnections between chromatin, transcription and splicing. Their relationship was mainly estimated through the direct crosstalk of machineries at specific locus (**Figure 6, black arrows**). However, their interconnection has to be also considered onto feed back loops that affect expression of factors composing these machineries. Here, our results demonstrate the essential role of splicing to regulate SUV39H2 and G9A expression, and this process impacts SUV39H2 function in transcription (**Figure 6, red arrows**). Our work illustrates the requirement to take into account this pathway to study chromatin factor function.

BIBLIOGRAPHIE

- Aggaard, L., Laible, G., Selenko, P., Schmid, M., Dorn, R., Schotta, G., Kuhfittig, S., Wolf, A., Lebersorger, A., Singh, P.B., et al. (1999). Functional mammalian homologues of the *Drosophila* PEV-modifier Su(var)3-9 encode centromere-associated proteins which complex with the heterochromatin component M31. *EMBO J.* *18*, 1923–1938.
- Aggaard, L., Schmid, M., Warburton, P., and Jenuwein, T. (2000). Mitotic phosphorylation of SUV39H1, a novel component of active centromeres, coincides with transient accumulation at mammalian centromeres. *J. Cell Sci.* *113* (Pt 5), 817–829.
- Abelson, J., Trotta, C.R., and Li, H. (1998). tRNA Splicing. *J. Biol. Chem.* *273*, 12685–12688.
- Aleman, L.M., Doench, J., and Sharp, P.A. (2007). Comparison of siRNA-induced off-target RNA and protein effects. *RNA* *13*, 385–395.
- Allemand, E., Batsché, E., and Muchardt, C. (2008). Splicing, transcription, and chromatin: a ménage à trois. *Curr. Opin. Genet. Dev.* *18*, 145–151.
- Alló, M., Buggiano, V., Fededa, J.P., Petrillo, E., Schor, I., de la Mata, M., Agirre, E., Plass, M., Eyra, E., Elela, S.A., et al. (2009). Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nat. Struct. Mol. Biol.* *16*, 717–724.
- De Almeida, S.F., Grosso, A.R., Koch, F., Fenouil, R., Carvalho, S., Andrade, J., Levezinho, H., Gut, M., Eick, D., Gut, I., et al. (2011). Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. *Nat. Struct. Mol. Biol.* *18*, 977–983.
- De Almeida, S.F., Grosso, A.R., Koch, F., Fenouil, R., Carvalho, S., Andrade, J., Levezinho, H., Gut, M., Eick, D., and Gut, I. (2011). Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. *Nat. Struct. Mol. Biol.* *18*, 977–983.
- Amara, S.G., Jonas, V., Rosenfeld, M.G., Ong, E.S., and Evans, R.M. (1982). Alternative RNA processing in calcitonin gene expression generates mRNAs encoding different polypeptide products. *Nature* *298*, 240–244.
- Ameur, A., Zaghlool, A., Halvardson, J., Wetterbom, A., Gyllenstein, U., Cavelier, L., and Feuk, L. (2011). Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat. Struct. Mol. Biol.* *18*, 1435–1440.
- Ameyar-Zazoua, M., Rachez, C., Souidi, M., Robin, P., Fritsch, L., Young, R., Morozova, N., Fenouil, R., Descostes, N., Andrau, J.-C., et al. (2012). Argonaute proteins couple chromatin silencing to alternative splicing. *Nat. Struct. Mol. Biol.* *19*, 998–1004.
- Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., and Zoghbi, H.Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* *23*, 185–188.
- Ancelin, K., Lange, U.C., Hajkova, P., Schneider, R., Bannister, A.J., Kouzarides, T., and Surani, M.A. (2006). Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells. *Nat. Cell Biol.* *8*, 623–630.
- Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C., and Komorowski, J. (2009). Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res.* *19*, 1732–1741.
- Auboeuf, D., Hönig, A., Berget, S.M., and O'Malley, B.W. (2002). Coordinate regulation of transcription and splicing by steroid receptor coregulators. *Science* *298*, 416–419.
- Ausió, J. (2006). Histone variants--the structure behind the function. *Brief. Funct. Genomic. Proteomic.* *5*, 228–243.

- Avvakumov, G. V, Walker, J.R., Xue, S., Li, Y., Duan, S., Bronner, C., Arrowsmith, C.H., and Dhe-Paganon, S. (2008). Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature* 455, 822–825.
- Ball, M.P., Li, J.B., Gao, Y., Lee, J.-H., LeProust, E.M., Park, I.-H., Xie, B., Daley, G.Q., and Church, G.M. (2009). Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat. Biotechnol.* 27, 361–368.
- Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res.* 21, 381–395.
- Batsché, E., Yaniv, M., and Muchardt, C. (2006). The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat. Struct. Mol. Biol.* 13, 22–29.
- Bell, G.I., Pictet, R.L., Rutter, W.J., Cordell, B., Tischler, E., and Goodman, H.M. (1980). Sequence of the human insulin gene. *Nature* 284, 26–32.
- Bell, M. V, Cowper, A.E., Lefranc, M., Bell, J.I., and Screaton, G.R. (1998). Influence of Intron Length on Alternative Splicing of CD44 Influence of Intron Length on Alternative Splicing of CD44. 18.
- Bell, O., Wirbelauer, C., Hild, M., Scharf, A.N.D., Schwaiger, M., MacAlpine, D.M., Zilbermann, F., van Leeuwen, F., Bell, S.P., Imhof, A., et al. (2007). Localized H3K36 methylation states define histone H4K16 acetylation during transcriptional elongation in *Drosophila*. *EMBO J.* 26, 4974–4984.
- Berget, S.M., Moore, C., and Sharp, P.A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA *. *Proc. Natl. Acad. Sci. U. S. A.* 74, 3171–3175.
- Berk, A.J., and Sharp, P.A. (1978). Spliced early mRNAs of simian virus 40. *Proc. Natl. Acad. Sci. U. S. A.* 75, 1274–1278.
- Bestor, T.H. (2000). The DNA methyltransferases of mammals. *Hum. Mol. Genet.* 9, 2395–2402.
- Bestor, T., Laudano, a, Mattaliano, R., and Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J. Mol. Biol.* 203, 971–983.
- Beyer, a L., and Osheim, Y.N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* 2, 754–765.
- Bezzi, M., Teo, S.X., Muller, J., Mok, W.C., Sahu, S.K., Vardy, L. a, Bonday, Z.Q., and Guccione, E. (2013). Regulation of constitutive and alternative splicing by PRMT5 reveals a role for Mdm4 pre-mRNA in sensing defects in the spliceosomal machinery. *Genes Dev.* 27, 1903–1916.
- Bhatt, D.M., Pandya-Jones, A., Tong, A.-J., Barozzi, I., Lissner, M.M., Natoli, G., Black, D.L., and Smale, S.T. (2012). Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* 150, 279–290.
- Bird, A. (1980). DNA methylation and the frequency of CpG in animal DNA Nucleic. *Nucleic Acids Res.* 8, 1499–1504.
- Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6–21.
- Boireau, S., Maiuri, P., Basyuk, E., de la Mata, M., Knezevich, A., Pradet-Balade, B., Bäcker, V., Kornblihtt, A.R., Marcello, A., and Bertrand, E. (2007). The transcriptional cycle of HIV-1 in real-time and live cells. *J. Cell Biol.* 179, 291–304.

Braunschweig, U., Gueroussov, S., Plocik, A.M., Graveley, B.R., and Blencowe, B.J. (2013). Dynamic integration of splicing within gene regulatory pathways. *Cell* 152, 1252–1269.

Breathnach, R., Mandel, J.L., and Chambon, P. (1977). Ovalbumin gene is split in chicken DNA. *Nature* 270, 314–319.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., Chambon, P., and Hare, K.O. (1978). Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proc. Natl. Acad. Sci. U. S. A.* 75, 4853–4857.

Bressan, G.C., Moraes, E.C., Manfiolli, A.O., Kuniyoshi, T.M., Passos, D.O., Gomes, M.D., and Kobarg, J. (2009). Arginine methylation analysis of the splicing-associated SR protein SFRS9/SRP30C. *Cell. Mol. Biol. Lett.* 14, 657–669.

Brett, D., Hanke, J., Lehmann, G., Haase, S., Delbrück, S., Krueger, S., Reich, J., and Bork, P. (2000). EST comparison indicates 38% of human mRNAs contain possible alternative splice forms. *FEBS Lett.* 474, 83–86.

Buljan, M., Chalancon, G., Dunker, a K., Bateman, A., Balaji, S., Fuxreiter, M., and Babu, M.M. (2013). Alternative splicing of intrinsically disordered regions and rewiring of protein interactions. *Curr. Opin. Struct. Biol.* 23, 443–450.

Buschhausen, G., Wittig, B., Graessmann, M., and Graessmann, a (1987). Chromatin structure is required to block transcription of the methylated herpes simplex virus thymidine kinase gene. *Proc. Natl. Acad. Sci. U. S. A.* 84, 1177–1181.

Calarco, J. a, Superina, S., O'Hanlon, D., Gabut, M., Raj, B., Pan, Q., Skalska, U., Clarke, L., Gelinas, D., van der Kooy, D., et al. (2009). Regulation of vertebrate nervous system alternative splicing and development by an SR-related protein. *Cell* 138, 898–910.

Candido, E.P.M., Reeves, R., and Davie, J.R. (1978). Sodium Butyrate Cultured Cells Inhibits Histone Deacetylation in. *Cell* 14, 105–113.

Carrillo Oesterreich, F., Preibisch, S., and Neugebauer, K.M. (2010). Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. *Mol. Cell* 40, 571–581.

Carroll, D.O., Scherthan, H., Peters, A.H.F.M., Opravil, S., Haynes, A.R., Laible, G., Schmid, M., Lebersorger, A., Sattler, L., Mattei, M.G., et al. (2000). Isolation and Characterization of Suv39h2 , a Second Histone H3 Methyltransferase Gene That Displays Testis-Specific Expression Isolation and Characterization of Suv39h2 , a Second Histone H3 Methyltransferase Gene That Displays Testis-Specific Expression.

Cereghini, S., and Yaniv, M. (1984). Assembly of transfected DNA into chromatin: structural changes in the origin-promoter-enhancer region upon replication. *EMBO J.* 3, 1243–1253.

Chang, B., Chen, Y., Zhao, Y., and Bruick, R.K. (2007). JMJD6 is a histone arginine demethylase. *Science* 318, 444–447.

Chen, M., and Manley, J.L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.* 10, 741–754.

Cheng, D., Côté, J., Shaaban, S., and Bedford, M.T. (2007). The arginine methyltransferase CARM1 regulates the coupling of transcription and mRNA processing. *Mol. Cell* 25, 71–83.

Chodavarapu, R.K., Feng, S., Bernatavichute, Y. V, Chen, P.-Y., Stroud, H., Yu, Y., Hetzel, J. a, Kuo, F., Kim, J., Cokus, S.J., et al. (2010). Relationship between nucleosome positioning and DNA methylation. *Nature* 466, 388–392.

- Choi, J.K. (2010). Contrasting chromatin organization of CpG islands and exons in the human genome. *Genome Biol.* *11*, R70.
- Choudhary, C., Kumar, C., Gnäd, F., Nielsen, M.L., Rehman, M., Walther, T.C., Olsen, J. V, and Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* *325*, 834–840.
- Chow, L.T., Gelinas, R.E., and Roberts, R.J. (1977). An Amazing Sequence Arrangement at the 5' Ends of Adenovirus 2 RNA. *Cell* *12*, 1–8.
- Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* *469*, 368–373.
- Close, P., East, P., Dirac-Svejstrup, a B., Hartmann, H., Heron, M., Maslen, S., Chariot, A., Söding, J., Skehel, M., and Svejstrup, J.Q. (2012). DBIRD complex integrates alternative mRNA splicing with RNA polymerase II transcript elongation. *Nature* *484*, 386–389.
- De Conti, L., Baralle, M., and Buratti, E. (2013). Exon and intron definition in pre-mRNA splicing. *Wiley Interdiscip. Rev. RNA* *4*, 49–60.
- Cooke, C., Hans, H., and Alwine, J.C. (1999). Utilization of splicing elements and polyadenylation signal elements in the coupling of polyadenylation and last-intron removal. *Mol. Cell. Biol.* *19*, 4971–4979.
- Coulter, D., and Greenleaf, A. (1985). A mutation in the largest subunit of RNA polymerase II alters RNA chain elongation in vitro. *J. Biol. Chem.* *260*, 13190–13198.
- Cramer, P., Pesce, C.G., Baralle, F.E., and Kornblihtt, A.R. (1997). Functional association between promoter structure and transcript alternative splicing. *Proc. Natl. Acad. Sci. U. S. A.* *94*, 11456–11460.
- Cramer, P., Cáceres, J.F., Cazalla, D., Kadener, S., Muro, a F., Baralle, F.E., and Kornblihtt, A.R. (1999). Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. *Mol. Cell* *4*, 251–258.
- Crawford, L. V, Cole, C.N., Smith, A.E., Paucha, E., Tegtmeyer, P., Rundell, K., and Berg, P. (1978). Organization and expression of early genes of simian virus 40. *Proc. Natl. Acad. Sci. U. S. A.* *75*, 117–121.
- Crick, F. (1979). Split Genes and RNA Splicing. *Science* *204*.
- Daniel, J.M., Spring, C.M., Crawford, H.C., Reynolds, A.B., and Baig, A. (2002). The p120(ctn)-binding partner Kaiso is a bi-modal DNA-binding protein that recognizes both a sequence-specific consensus and methylated CpG dinucleotides. *Nucleic Acids Res.* *30*, 2911–2919.
- Darnell, J.E., Lodish, H.F., and Baltimore, D. (1986). *Molecular Cell Biology* (New York: Scientific American Books).
- Das, R., Zhou, Z., and Reed, R. (2000). Functional Association of U2 snRNP with the ATP-Independent Spliceosomal Complex E. *Mol. Cell* *5*, 779–787.
- Das, R., Dufu, K., Romney, B., Feldt, M., Elenko, M., and Reed, R. (2006). Functional coupling of RNAP II transcription to spliceosome assembly. *Genes Dev.* *20*, 1100–1109.
- Das, R., Yu, J., Zhang, Z., Gygi, M.P., Krainer, A.R., Gygi, S.P., and Reed, R. (2007). SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol. Cell* *26*, 867–881.
- Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. *Genes Dev.* *25*, 1010–1022.

- Defossez, P.-A., and Stancheva, I. (2011). Biological functions of methyl-CpG-binding proteins. (Elsevier Inc.).
- Dhami, P., Saffrey, P., Bruce, A.W., Dillon, S.C., Chiang, K., Bonhoure, N., Koch, C.M., Bye, J., James, K., Foad, N.S., et al. (2010). Complex exon-intron marking by histone modifications is not determined solely by nucleosome distribution. *PLoS One* 5, e12339.
- Ding, S., Shi, J., Qian, W., Iqbal, K., Grundke-Iqbal, I., Gong, C.-X., and Liu, F. (2012). Regulation of alternative splicing of tau exon 10 by 9G8 and Dyrk1A. *Neurobiol. Aging* 33, 1389–1399.
- Dowhan, D.H., Hong, E.P., Auboeuf, D., Dennis, A.P., Wilson, M.M., Berget, S.M., and O'Malley, B.W. (2005). Steroid hormone receptor coactivation and alternative RNA splicing by U2AF65-related proteins CAPERalpha and CAPERbeta. *Mol. Cell* 17, 429–439.
- Dujardin, G., Lafaille, C., and Petrillo, E. (2013). Transcriptional elongation and alternative splicing. ... *Biophys. Acta (BBA)* ...
- Durdevic, Z., and Schaefer, M. (2013). Dnmt2 methyltransferases and immunity: an ancient overlooked connection between nucleotide modification and host defense? *Bioessays* 35, 1044–1049.
- Ellis, J.D., Barrios-rodiles, M., Recep, C., Irimia, M., Kim, T., Calarco, J.A., Wang, X., Pan, Q., Hanlon, D.O., Kim, P.M., et al. (2012). Tissue-Specific Alternative Splicing Remodels Protein-Protein Interaction Networks. *Mol. Cell* 46, 884–892.
- Van Emburgh, B.O., and Robertson, K.D. (2011). Modulation of Dnmt3b function in vitro by interactions with Dnmt3L, Dnmt3a and Dnmt3b splice variants. *Nucleic Acids Res.* 39, 4984–5002.
- Estève, P.-O., Patnaik, D., Chin, H.G., Benner, J., Teitell, M. a, and Pradhan, S. (2005). Functional analysis of the N- and C-terminus of mammalian G9a histone H3 methyltransferase. *Nucleic Acids Res.* 33, 3211–3223.
- Estève, P.-O., Chin, H.G., Smallwood, A., Feehery, G.R., Gangisetty, O., Karpf, A.R., Carey, M.F., and Pradhan, S. (2006). Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes Dev.* 20, 3089–3103.
- Feng, S., Cokus, S.J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E., et al. (2010). Conservation and divergence of methylation patterning in plants and animals. *Proc. Natl. Acad. Sci. U. S. A.* 107, 8689–8694.
- Fica, S.M., Tuttle, N., Novak, T., Li, N.-S., Lu, J., Koodathingal, P., Dai, Q., Staley, J.P., and Piccirilli, J. a (2013). RNA catalyses nuclear pre-mRNA splicing. *Nature* 503, 229–234.
- Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T. a, Marques, C.J., Andrews, S., and Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* 473, 398–402.
- Filion, G.J.P., Zhenilo, S., Salozhin, S., Yamada, D., Prokhortchouk, E., and Defossez, P. (2006). A Family of Human Zinc Finger Proteins That Bind Methylated DNA and Repress Transcription A Family of Human Zinc Finger Proteins That Bind Methylated DNA and Repress Transcription.
- Fox-Walsh, K.L., Dou, Y., Lam, B.J., Hung, S.-P., Baldi, P.F., and Hertel, K.J. (2005). The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proc. Natl. Acad. Sci. U. S. A.* 102, 16176–16181.
- Friesen, W.J., Paushkin, S., Wyce, A., Massenet, S., Pesiridis, G.S., Duyne, G. Van, Rappsilber, J., Mann, M., and Dreyfuss, G. (2001). The Methylosome , a 20S Complex Containing JBP1 and pICln , Produces Dimethylarginine-Modified Sm Proteins The Methylosome , a 20S Complex Containing JBP1 and pICln , Produces Dimethylarginine-Modified Sm Proteins.

Fritsch, L., Robin, P., Mathieu, J.R.R., Souidi, M., Hinaux, H., Rougeulle, C., Harel-bellan, A., Ameyar-Zazoua, M., and Ait-si-ali, S. (2010). A subset of the histone H3 lysine 9 methyltransferases Suv39h1, G9a, GLP, and SETDB1 participate in a multimeric complex. *Mol. Cell* 37, 46–56.

Fuks, F., Hurd, P.J., Deplus, R., and Kouzarides, T. (2003). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Res.* 31, 2305–2312.

Gallwitz, D., and Sures, I. (1980). Structure of a split yeast gene: complete nucleotide sequence of the actin gene in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* 77, 2546–2550.

García-Cao, M., O’Sullivan, R., Peters, A.H.F.M., Jenuwein, T., and Blasco, M.A. (2004). Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases. *Nat. Genet.* 36, 94–99.

Ge, H., Si, Y., and Wolffe, a P. (1998). A novel transcriptional coactivator, p52, functionally interacts with the essential splicing factor ASF/SF2. *Mol. Cell* 2, 751–759.

Geiman, T.M., Sankpal, U.T., Robertson, A.K., Zhao, Y., Zhao, Y., and Robertson, K.D. (2004). DNMT3B interacts with hSNF2H chromatin remodeling enzyme, HDACs 1 and 2, and components of the histone methylation system. *Biochem. Biophys. Res. Commun.* 318, 544–555.

Gilbert, W. (1978). Why genes in pieces? *Nature* 271, 501.

Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* 74, 481–514.

Goll, M.G., Kirpekar, F., Maggert, K. a, Yoder, J. a, Hsieh, C.-L., Zhang, X., Golic, K.G., Jacobsen, S.E., and Bestor, T.H. (2006). Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2. *Science* 311, 395–398.

Gopalakrishnan, S., Emburgh, B.O. Van, Shan, J., Su, Z., Fields, C.R., Vieweg, J., Hamazaki, T., Schwartz, P.H., Terada, N., and Robertson, K.D. (2009). A Novel DNMT3B Splice Variant Expressed in Tumor and Pluripotent Cells Modulates Genomic DNA Methylation Patterns and Displays Altered DNA Binding. *Mol. Cancer Res.* 7, 1622–1634.

Görnemann, J., Kotovic, K.M., Hujer, K., and Neugebauer, K.M. (2005). Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. *Mol. Cell* 19, 53–63.

De Graaf, K., Czajkowska, H., Rottmann, S., Packman, L.C., Lilischkis, R., Lüscher, B., and Becker, W. (2006). The protein kinase DYRK1A phosphorylates the splicing factor SF3b1/SAP155 at Thr434, a novel in vivo phosphorylation site. *BMC Biochem.* 7, 7.

Grainger, R.J., and Beggs, J.D. (2005). Prp8 protein : At the heart of the spliceosome Prp8 protein : At the heart of the spliceosome. 533–557.

Green, M.R., Maniatis, T., and Melton, D.A. (1983). Human beta-globin pre-mRNA synthesized in vitro is accurately spliced in *Xenopus* oocyte nuclei. *Cell* 32, 681–694.

Gregory, B.L., and Cheung, V.G. (2014). Natural variation in the histone demethylase, KDM4C, influences expression levels of specific genes including those that affect cell growth. *Genome Res.* 24, 52–63.

Grewal, S.I. (2010). RNAi-dependent formation of heterochromatin and its diverse functions. *Curr. Opin. Genet. Dev.* 20, 134–141.

Grover, A., Houlden, H., Baker, M., Adamson, J., Lewis, J., Prihar, G., Pickering-Brown, S., Duff, K., and Hutton, M. (1999). 5’ Splice Site Mutations in tau Associated with the Inherited Dementia FTDP-17 Affect a Stem-Loop Structure That Regulates Alternative Splicing of Exon 10. *J. Biol. Chem.* 274, 15134–15143.

- Grzenda, A., Lomber, G., Svingen, P., Mathison, A., Calvo, E., Iovanna, J., Xiong, Y., Faubion, W., and Urrutia, R. (2013). Functional characterization of EZH2beta reveals the increased complexity of EZH2 isoforms involved in the regulation of mammalian gene expression. *Epigenetics Chromatin* 6, 3.
- Hagemann, S., Heil, O., Lyko, F., and Brueckner, B. (2011). Azacytidine and decitabine induce gene-specific and non-random DNA demethylation in human cancer cell lines. *PLoS One* 6, e17388.
- Han, H., Irimia, M., Ross, P.J., Sung, H.-K., Alipanahi, B., David, L., Golipour, A., Gabut, M., Michael, I.P., Nachman, E.N., et al. (2013). MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. *Nature* 498, 241–245.
- Han, J., Kim, D., and Morris, K. V (2007). Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells. *Proc. Natl. Acad. Sci. U. S. A.* 104, 12422–12427.
- Hastings, M.L., Allemand, E., Duelli, D.M., Myers, M.P., and Krainer, A.R. (2007). Control of pre-mRNA splicing by the general splicing factors PUF60 and U2AF(65). *PLoS One* 2, e538.
- He, Y., Li, B., Li, Z., Liu, P., Wang, Y., Tang, Q., and Ding, J. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* (80-.). 333, 1303–1307.
- Hellman, A., and Chess, A. (2007). Gene body-specific methylation on the active X chromosome. *Science* 315, 1141–1143.
- Hendrich, B., and Bird, A. (1998). Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins. *Mol. Cell. Biol.* 18, 6538–6547.
- Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V. a, and Bird, a (2001). Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes Dev.* 15, 710–723.
- Hertel, K.J., and Maniatis, T. (1999). Serine-arginine (SR)-rich splicing factors have an exon-independent function in pre-mRNA splicing. *Proc. Natl. Acad. Sci. U. S. A.* 96, 2651–2655.
- Hicks, M.J., Yang, C.-R., Kotlajich, M. V, and Hertel, K.J. (2006). Linking splicing to Pol II transcription stabilizes pre-mRNAs and influences splicing patterns. *PLoS Biol.* 4, e147.
- Hiller, M., Zhang, Z., Backofen, R., and Stamm, S. (2007). Pre-mRNA secondary structures influence exon recognition. *PLoS Genet.* 3, e204.
- Hirose, Y., Tacke, R., and Manley, J.L. (1999). Phosphorylated RNA polymerase II stimulates pre-mRNA splicing polymerase II stimulates pre-mRNA splicing. 1234–1239.
- Hnilicová, J., Hozeifi, S., Dušková, E., Icha, J., Tománková, T., and Staněk, D. (2011). Histone deacetylase activity modulates alternative splicing. *PLoS One* 6, e16727.
- Hodges, C., Bintu, L., Lubkowska, L., Kashlev, M., and Bustamante, C. (2009). Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science* 325, 626–628.
- Hon, G., Wang, W., and Ren, B. (2009). Discovery and annotation of functional chromatin signatures in the human genome. *PLoS Comput. Biol.* 5, e1000566.
- Hopfner, R., Mousli, M., Jeltsch, J., Lutz, Y., Marin, C., Bellocq, J., Oudet, P., and Bronner, C. (2000). ICBP90 , a Novel Human CCAAT Binding Protein , Involved in the Regulation of Topoisomerase II α Expression ICBP90 , a Novel Human CCAAT Binding Protein , Involved in the Regulation of. 121–128.
- House, A.E., and Lynch, K.W. (2006). An exonic splicing silencer represses spliceosome assembly after ATP-dependent exon recognition. *Nat. Struct. Mol. Biol.* 13, 937–944.

- Howe, K.J., Kane, C.M., and Ares, M. (2003). Perturbation of transcription elongation influences the fidelity of internal exon inclusion in *Saccharomyces cerevisiae*. *RNA* 9, 993–1006.
- Hsin, J.-P., and Manley, J.L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev.* 26, 2119–2137.
- Hsu, J.Y., Sun, Z.W., Li, X., Reuben, M., Tatchell, K., Bishop, D.K., Grushcow, J.M., Brame, C.J., Caldwell, J. a, Hunt, D.F., et al. (2000). Mitotic phosphorylation of histone H3 is governed by Ipl1/aurora kinase and Glc7/PP1 phosphatase in budding yeast and nematodes. *Cell* 102, 279–291.
- Huang, Y., Li, W., Yao, X., Lin, Q.-J., Yin, J.-W., Liang, Y., Heiner, M., Tian, B., Hui, J., and Wang, G. (2012). Mediator complex regulates alternative mRNA processing via the MED23 subunit. *Mol. Cell* 45, 459–469.
- Huff, J.T., Plocik, A.M., Guthrie, C., and Yamamoto, K.R. (2010). Reciprocal intronic and exonic histone modification regions in humans. *Nat. Struct. Mol. Biol.* 17, 1495–1499.
- Ibrahim, E.C., Schaal, T.D., Hertel, K.J., Reed, R., and Maniatis, T. (2005). Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers. *Proc. Natl. Acad. Sci. U. S. A.* 102, 5002–5007.
- Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R.W., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R., and Bird, A.P. (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet.* 6, e1001134.
- Ip, J.Y., Schmidt, D., Pan, Q., Ramani, A.K., Fraser, A.G., Odom, D.T., and Blencowe, B.J. (2011). Global impact of RNA polymerase II elongation inhibition on alternative splicing regulation. *Genome Res.* 21, 390–401.
- Irimia, M., and Blencowe, B.J. (2012). Alternative splicing: decoding an expansive regulatory layer. *Curr. Opin. Cell Biol.* 24, 323–332.
- Ito, S., D'Alessio, A.C., Taranova, O. V, Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 466, 1129–1133.
- Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J. a, He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333, 1300–1303.
- Izban, M.G., and Luse, D.S. (1991). Transcription on nucleosomal templates by RNA polymerase II in vitro: inhibition of elongation with enhancement of sequence-specific pausing. *Genes Dev.* 5, 683–696.
- Jackson, A.L., Burchard, J., Leake, D., Reynolds, A., Schelter, J., Guo, J.I.E., Johnson, J.M., Lim, L.E.E., Karpilow, J.O.N., Nichols, K.I.M., et al. (2006). Position-specific chemical modification of siRNAs reduces “off-target” transcript silencing. 1197–1205.
- Jair, K.-W., Bachman, K.E., Suzuki, H., Ting, A.H., Rhee, I., Yen, R.-W.C., Baylin, S.B., and Schuebel, K.E. (2006). De novo CpG island methylation in human cancer cells. *Cancer Res.* 66, 682–692.
- Jeffreys, A.J., and Flavell, R.A. (1977). The Rabbit beta Globin Gene Contains a Large Insert in the Coding Sequence. *Cell* 12, 1097–1108.
- Ji, X., Zhou, Y., Pandit, S., Huang, J., Li, H., Lin, C.Y., Xiao, R., Burge, C.B., and Fu, X.-D. (2013). SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. *Cell* 153, 855–868.
- Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A., and Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature* 449, 248–251.

Johnson, C. a, O'Neill, L.P., Mitchell, a, and Turner, B.M. (1998). Distinctive patterns of histone H4 acetylation are associated with defined sequence elements within both heterochromatic and euchromatic regions of the human genome. *Nucleic Acids Res.* 26, 994–1001.

Johnson, J.M., Castle, J., Garrett-Engle, P., Kan, Z., Loerch, P.M., Armour, C.D., Santos, R., Schadt, E.E., Stoughton, R., and Shoemaker, D.D. (2003). Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* 302, 2141–2144.

Jolla, L., Ng, R., and Abelson, J. (1980). Isolation and sequence of the gene for actin in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* 77, 3912–3916.

Jurica, M.S., and Moore, M.J. (2003). Pre-mRNA Splicing: Awash in a Sea of Proteins. *Mol. Cell* 12, 5–14.

Kadener, S., Cramer, P., Nogués, G., Cazalla, D., de la Mata, M., Fededa, J.P., Werbajh, S.E., Srebrow, A., and Kornblihtt, A.R. (2001). Antagonistic effects of T-Ag and VP16 reveal a role for RNA pol II elongation on alternative splicing. *EMBO J.* 20, 5759–5768.

Kadener, S., Fededa, J.P., Rosbash, M., and Kornblihtt, A.R. (2002). Regulation of alternative splicing by a transcriptional enhancer through RNA pol II elongation. *Proc. Natl. Acad. Sci. U. S. A.* 99, 8185–8190.

Kanopka, A., Mühlemann, O., and Akusjärvi, G. (1996). Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature* 381.

Karimi, M.M., Goyal, P., Maksakova, I. a, Bilenky, M., Leung, D., Tang, J.X., Shinkai, Y., Mager, D.L., Jones, S., Hirst, M., et al. (2011). DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. *Cell Stem Cell* 8, 676–687.

Kass, S.U., Landsberger, N., and Wolffe, a P. (1997). DNA methylation directs a time-dependent repression of transcription initiation. *Curr. Biol.* 7, 157–165.

Keren-Shaul, H., Lev-Maor, G., and Ast, G. (2013). Pre-mRNA splicing is a determinant of nucleosome organization. *PLoS One* 8, e53506.

Khan, D.H., Gonzalez, C., Cooper, C., Sun, J.-M., Chen, H.Y., Healy, S., Xu, W., Smith, K.T., Workman, J.L., Leygue, E., et al. (2013). RNA-dependent dynamic histone acetylation regulates MCL1 alternative splicing. *Nucleic Acids Res.* 1–15.

Khodor, Y.L., Rodriguez, J., Abruzzi, K.C., Tang, C.-H.A., Marr, M.T., and Rosbash, M. (2011). Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in *Drosophila*. *Genes Dev.* 25, 2502–2512.

Khorasanizadeh, S. (2004). The Nucleosome : From Genomic Organization to Genomic Regulation University of Virginia Health System. 116, 259–272.

Kim, D.H., Villeneuve, L.M., Morris, K. V, and Rossi, J.J. (2006). Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nat. Struct. Mol. Biol.* 13, 793–797.

Kim, J.K., Estève, P.-O., Jacobsen, S.E., and Pradhan, S. (2009). UHRF1 binds G9a and participates in p21 transcriptional regulation in mammalian cells. *Nucleic Acids Res.* 37, 493–505.

Kim, S., Kim, H., Fong, N., Erickson, B., and Bentley, D.L. (2011a). Pre-mRNA splicing is a determinant of histone H3K36 methylation. *Proc. Natl. Acad. Sci. U. S. A.* 108, 13564–13569.

Kim, S., Kim, H., Fong, N., Erickson, B., and Bentley, D.L. (2011b). Pre-mRNA splicing is a determinant of histone H3K36 methylation. *Proc. Natl. Acad. Sci.* 108, 13564–13569.

Knoop, V. (2011). When you can't trust the DNA: RNA editing changes transcript sequences. *Cell. Mol. Life Sci.* 68, 567–586.

- Kohli, R.M., and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* 502, 472–479.
- Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X.S., and Ahringer, J. (2009). Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.* 41, 376–381.
- Konarska, M.M., Padgett, R. a, and Sharp, P. a (1984). Recognition of cap structure in splicing in vitro of mRNA precursors. *Cell* 38, 731–736.
- Kondo, E., Gu, Z., Horii, A., and Fukushima, S. (2005). The Thymine DNA Glycosylase MBD4 Represses Transcription and Is Associated with Methylated p16 INK4a and hMLH1 Genes. *Mol. Cell. Biol.* 25, 4388–4396.
- Kornblihtt, A.R., Schor, I.E., Alló, M., Dujardin, G., Petrillo, E., and Muñoz, M.J. (2013). Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* 14, 153–165.
- Kouzarides, T. (2007a). Chromatin modifications and their function. *Cell* 128, 693–705.
- Kouzarides, T. (2007b). SnapShot: Histone-modifying enzymes. *Cell* 131, 822.
- Krainer, A.R., Maniatis, T., Ruskin, B., and Green, M.R. (1984). Normal and mutant human beta-globin pre-mRNAs are faithfully and efficiently spliced in vitro. *Cell* 36, 993–1005.
- Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324, 929–930.
- Kuhn, A.N., van Santen, M. a, Schwienhorst, A., Urlaub, H., and Lührmann, R. (2009). Stalling of spliceosome assembly at distinct stages by small-molecule inhibitors of protein acetylation and deacetylation. *RNA* 15, 153–175.
- Kyburz, A., Friedlein, A., Langen, H., and Keller, W. (2006). Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing. *Mol. Cell* 23, 195–205.
- De la Mata, M., Alonso, C.R., Fededa, J.P., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A Slow RNA Polymerase II Affects Alternative Splicing In Vivo. *Mol. Cell* 12, 525–532.
- De la Mata, M., Lafaille, C., and Kornblihtt, A.R. (2010). First come, first served revisited: factors affecting the same alternative splicing event have different effects on the relative rates of intron removal. *RNA* 16, 904–912.
- Lacadie, S. a, and Rosbash, M. (2005). Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5' ss base pairing in yeast. *Mol. Cell* 19, 65–75.
- Lacroix, M., El Messaoudi, S., Rodier, G., Le Cam, A., Sardet, C., and Fabrizio, E. (2008). The histone-binding protein COPR5 is required for nuclear functions of the protein arginine methyltransferase PRMT5. *EMBO Rep.* 9, 452–458.
- Lambowitz, A.M., and Zimmerly, S. (2011). Group II introns: mobile ribozymes that invade DNA. *Cold Spring Harb. Perspect. Biol.* 3, a003616.
- Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C.T., Low, H.M., Kin Sung, K.W., Rigoutsos, I., Loring, J., et al. (2010). Dynamic changes in the human methylome during differentiation. *Genome Res.* 20, 320–331.
- Lee, M.G., Wynder, C., Cooch, N., and Shiekhata, R. (2005). An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. *Nature* 437, 432–435.

- Leff, S.E., and Rosenfeld, M.G. (1986). Complex transcriptional units: Diversity in gene expression by alternative sRNA processing. *Annu. Rev. Biochem.* *55*, 1091–1117.
- Lei, H., Oh, S.P., Okano, M., Jüttermann, R., Goss, K. a, Jaenisch, R., and Li, E. (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* *122*, 3195–3205.
- Lewis, J.D., Izaurralde, E., Jarmolowski, a, McGuigan, C., and Mattaj, I.W. (1996). A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes Dev.* *10*, 1683–1698.
- Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* *69*, 915–926.
- Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S., and Fu, X.-D. (2008). The splicing factor SC35 has an active role in transcriptional elongation. *Nat. Struct. Mol. Biol.* *15*, 819–826.
- Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* *462*, 315–322.
- Listerman, I., Sapra, A.K., and Neugebauer, K.M. (2006). Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nat. Struct. Mol. Biol.* *13*, 815–822.
- Lopez-Atalaya, J.P., Ito, S., Valor, L.M., Benito, E., and Barco, A. (2013). Genomic targets, and histone acetylation and gene expression profiling of neural HDAC inhibition. *Nucleic Acids Res.* *41*, 8072–8084.
- Lorincz, M.C., Dickerson, D.R., Schmitt, M., and Groudine, M. (2004). Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nat. Struct. Mol. Biol.* *11*, 1068–1075.
- Loyola, a, LeRoy, G., Wang, Y.H., and Reinberg, D. (2001). Reconstitution of recombinant chromatin establishes a requirement for histone-tail modifications during chromatin assembly and transcription. *Genes Dev.* *15*, 2837–2851.
- Luco, R., Alló, M., Schor, I., Kornblihtt, A., and Misteli, T. (2011). Epigenetics in alternative pre-mRNA splicing. *Cell* *144*, 16–26.
- Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M., and Misteli, T. (2010). Regulation of alternative splicing by histone modifications. *Science* *327*, 996–1000.
- Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C., and Maleszka, R. (2010). The honey bee epigenomes: differential methylation of brain DNA in queens and workers. *PLoS Biol.* *8*, e1000506.
- Lynch, C.J., Shah, Z.H., Allison, S.J., Ahmed, S.U., Ford, J., Warnock, L.J., Li, H., Serrano, M., and Milner, J. (2010). SIRT1 undergoes alternative splicing in a novel auto-regulatory loop with p53. *PLoS One* *5*, e13502.
- Ma, H., Baumann, C., Li, H., and Strahl, B. (2001). Hormone-dependent, CARM1-directed, arginine-specific methylation of histone H3 on a steroid-regulated promoter. *Curr. Biol.* *11*, 1981–1985.
- Martin, R.M., Rino, J., Carvalho, C., Kirchhausen, T., and Carmo-Fonseca, M. (2013). Live-cell visualization of pre-mRNA splicing with single-molecule sensitivity. *Cell Rep.* *4*, 1144–1155.
- Martín Caballero, I., Hansen, J., Leaford, D., Pollard, S., and Hendrich, B.D. (2009). The methyl-CpG binding proteins MeCP2, Mbd2 and Kaiso are dispensable for mouse embryogenesis, but play a redundant function in neural differentiation. *PLoS One* *4*, e4315.
- Matsuda, H., Paul, B.D., Choi, C.Y., and Shi, Y.-B. (2007). Contrasting Effects of Two Alternative Splicing Forms of Coactivator-Associated Arginine Methyltransferase 1 on Thyroid Hormone Receptor-Mediated Transcription in *Xenopus laevis*. *Mol. Endocrinol.* *21*, 1082–1094.

- Matsuo, K., Silke, J., Gramatikoff, K., and Schaffner, W. (1994). The CpG-specific methylase SssI has topoisomerase activity in the presence of Mg²⁺. *Nucleic Acids Res.* 22, 5354–5359.
- Matter, N., Herrlich, P., and König, H. (2002). Signal-dependent regulation of splicing via phosphorylation of Sam68. *Nature* 420, 691–695.
- Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D’Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y., et al. (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* 466, 253–257.
- Maunakea, A.K., Chepelev, I., Cui, K., and Zhao, K. (2013). Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res.* 23, 1256–1269.
- Mayeda, A., Helfman, D., and Krainer, A. (1993). Modulation of exon skipping and inclusion by heterogeneous nuclear ribonucleoprotein A1 and pre-mRNA splicing factor SF2/ASF. *Mol. Cell.* ... 13.
- McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S., Wilkens, M., and Bentley, D. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature*.
- Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L., and Bird, a P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* 58, 499–507.
- Meister, G., Eggert, C., Bühler, D., Brahms, H., Kambach, C., and Fischer, U. (2001). Methylation of Sm proteins by a complex containing PRMT5 and the putative U snRNP assembly factor pICln. *Curr. Biol.* 11, 1990–1994.
- Melcher, M., Schmid, M., Aagaard, L., Selenko, P., Laible, G., and Jenuwein, T. (2000). Structure-function analysis of SUV39H1 reveals a dominant role in heterochromatin organization, chromosome segregation, and mitotic progression. *Mol. Cell. Biol.* 20, 3728–3741.
- Metzger, E., Wissmann, M., Yin, N., Müller, J.M., Schneider, R., Peters, A.H.F.M., Günther, T., Buettner, R., and Schüle, R. (2005). LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* 437, 436–439.
- Militello, K.T., Simon, R.D., Qureshi, M., Maines, R., VanHorne, M.L., Hennick, S.M., Jayakar, S.K., and Pounder, S. (2012). Conservation of Dcm-mediated cytosine DNA methylation in *Escherichia coli*. *FEMS Microbiol. Lett.* 328, 78–85.
- Mironov, A.A., Wildon Fickett, J., and Gelfand, M.S. (1999). Frequent Alternative Splicing of Human Genes. *Genome Res.* 9, 1288–1293.
- Misteli, T., Cáceres, J., and Spector, D. (1997). The dynamics of a pre-mRNA splicing factor in living cells. *Nature* 387, 523–527.
- Monnier, P., Martinet, C., Pontis, J., Stancheva, I., Ait-Si-Ali, S., and Dandolo, L. (2013). H19 lncRNA controls gene expression of the Imprinted Gene Network by recruiting MBD1. *Proc. Natl. Acad. Sci. U. S. A.* 110, 20693–20698.
- Morris, K. V, Chan, S.W.-L., Jacobsen, S.E., and Looney, D.J. (2004). Small interfering RNA-induced transcriptional gene silencing in human cells. *Science* 305, 1289–1292.
- Mossman, D., Kim, K.-T., and Scott, R.J. (2010). Demethylation by 5-aza-2'-deoxycytidine in colorectal cancer cells targets genomic DNA whilst promoter CpG island methylation persists. *BMC Cancer* 10, 366.

- Mozzetta, C., Pontis, J., Fritsch, L., Robin, P., Portoso, M., Proux, C., Margueron, R., and Ait-Si-Ali, S. (2013). The Histone H3 Lysine 9 Methyltransferases G9a and GLP Regulate Polycomb Repressive Complex 2-Mediated Gene Silencing. *Mol. Cell* 277–289.
- Muñoz, M.J., de la Mata, M., and Kornblihtt, A.R. (2010). The carboxy terminal domain of RNA polymerase II and alternative splicing. *Trends Biochem. Sci.* 35, 497–504.
- Musselman, C. a, Lalonde, M.-E., Côté, J., and Kutateladze, T.G. (2012). Perceiving the epigenetic landscape through histone readers. *Nat. Struct. Mol. Biol.* 19, 1218–1227.
- Nan, X., Ng, H.H., Johnson, C. a, Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, a (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 393, 386–389.
- Natalizio, B.J., Robson-Dixon, N.D., and Garcia-Blanco, M. a (2009). The Carboxyl-terminal Domain of RNA Polymerase II Is Not Sufficient to Enhance the Efficiency of Pre-mRNA Capping or Splicing in the Context of a Different Polymerase. *J. Biol. Chem.* 284, 8692–8702.
- Nativio, R., Sparago, A., Ito, Y., Weksberg, R., Riccio, A., and Murrell, A. (2011). Disruption of genomic neighbourhood at the imprinted IGF2-H19 locus in Beckwith-Wiedemann syndrome and Silver-Russell syndrome. *Hum. Mol. Genet.* 20, 1363–1374.
- Neugebauer, K.M., and Roth, M.B. (1997). Distribution of pre-mRNA splicing factors at sites of RNA polymerase II transcription. *Genes Dev.* 11, 1148–1159.
- Ng, H.H., Zhang, Y., Hendrich, B., Johnson, C. a, Turner, B.M., Erdjument-Bromage, H., Tempst, P., Reinberg, D., and Bird, a (1999). MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat. Genet.* 23, 58–61.
- Nielsen, H., and Johansen, S.D. (2009). Group I introns: Moving in new directions. *RNA Biol.* 6, 375–383.
- Niwa, M., and Berget, S.M. (1991). Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. *Genes Dev.* 5, 2086–2095.
- Nogués, G., Muñoz, M.J., and Kornblihtt, A.R. (2003). Influence of polymerase II processivity on alternative splicing depends on splice site strength. *J. Biol. Chem.* 278, 52166–52171.
- Noma, K., Sugiyama, T., Cam, H., Verdel, A., Zofall, M., Jia, S., Moazed, D., and Grewal, S.I.S. (2004). RITS acts in cis to promote RNA interference-mediated transcriptional and post-transcriptional silencing. *Nat. Genet.* 36, 1174–1180.
- Ohkura, N., Takahashi, M., Yaguchi, H., Nagamura, Y., and Tsukada, T. (2005). Coactivator-associated arginine methyltransferase 1, CARM1, affects pre-mRNA splicing in an isoform-specific manner. *J. Biol. Chem.* 280, 28927–28935.
- Okano, M., Bell, D.W., Haber, D. a, and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99, 247–257.
- Osheim, Y.N., Miller, O.L., and Beyer, a L. (1985). RNP particles at splice junction sequences on Drosophila chorion transcripts. *Cell* 43, 143–151.
- Pal, S., Yun, R., Datta, A., Lacomis, L., Erdjument-bromage, H., Kumar, J., Tempst, P., and Sif, S. (2003). mSin3A / Histone Deacetylase 2- and PRMT5-Containing Brg1 Complex Is Involved in Transcriptional Repression of the Myc Target Gene cad. 23, 7475–7487.
- Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40, 1413–1415.

- Pandya-Jones, A., and Black, D.L. (2009). Co-transcriptional splicing of constitutive and alternative exons. *RNA* *15*, 1896–1908.
- Pandya-Jones, A., Bhatt, D.M., Lin, C.-H., Tong, A.-J., Smale, S.T., and Black, D.L. (2013). Splicing kinetics and transcript release from the chromatin compartment limit the rate of Lipid A-induced gene expression. *RNA* *19*, 811–827.
- Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R., and Dodgson, J. (1980). The Evolution of Genes: the Chicken Preproinsulin Gene. *Cell* *20*, 555–566.
- Peters, A., O’Carroll, D., and Scherthan, H. (2001). Loss of the Suv39h Histone Methyltransferases Impairs Mammalian Heterochromatin and Genome Stability. *Cell* *107*, 323–337.
- Phizicky, E.M., and Hopper, A.K. (2010). tRNA biology charges to the front. *Genes Dev.* *24*, 1832–1860.
- Piotrowska, H., and Jagodzinski, P.P. (2009). Glucocorticoid receptor alpha and beta variant expression is associated with ASF/SF2 splicing factor upregulation in HT-29 colon cancer and MCF-7 breast carcinoma cells. *Arch. Med. Res.* *40*, 156–162.
- Pradeepa, M.M., Sutherland, H.G., Ule, J., Grimes, G.R., and Bickmore, W. a (2012). Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. *PLoS Genet.* *8*, e1002717.
- Prokhortchouk, a, Hendrich, B., Jørgensen, H., Ruzov, a, Wilm, M., Georgiev, G., Bird, a, and Prokhortchouk, E. (2001). The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes Dev.* *15*, 1613–1618.
- Pruitt, K., Zinn, R.L., Ohm, J.E., McGarvey, K.M., Kang, S.-H.L., Watkins, D.N., Herman, J.G., and Baylin, S.B. (2006). Inhibition of SIRT1 reactivates silenced cancer genes without loss of promoter DNA hypermethylation. *PLoS Genet.* *2*, e40.
- Qian, W., Liang, H., Shi, J., Jin, N., Grundke-Iqbal, I., Iqbal, K., Gong, C.-X., and Liu, F. (2011). Regulation of the alternative splicing of tau exon 10 by SC35 and Dyrk1A. *Nucleic Acids Res.* *39*, 6161–6171.
- Rathert, P., Dhayalan, A., Murakami, M., Zhang, X., Tamas, R., Jurkowska, R., Komatsu, Y., Shinkai, Y., Cheng, X., and Jeltsch, A. (2008). Protein lysine methyltransferase G9a acts on non-histone targets. *Nat. Chem. Biol.* *4*, 344–346.
- Rauch, T.A., Wu, X., Zhong, X., Riggs, A.D., and Pfeifer, G.P. (2008). A human B cell methylome at 100 % base pair resolution *BIOLOGY*.
- Reenan, R. a, Hanrahan, C.J., and Ganetzky, B. (2000). The mle(napts) RNA helicase mutation in drosophila results in a splicing catastrophe of the para Na⁺ channel transcript in a region of RNA editing. *Neuron* *25*, 139–149.
- Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* *447*, 425–432.
- Renda, M., Baglivo, I., Burgess-Beusse, B., Esposito, S., Fattorusso, R., Felsenfeld, G., and Pedone, P. V (2007). Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. *J. Biol. Chem.* *282*, 33336–33345.
- Rhee, I., Jair, K.W., Yen, R.W., Lengauer, C., Herman, J.G., Kinzler, K.W., Vogelstein, B., Baylin, S.B., and Schuebel, K.E. (2000). CpG methylation is maintained in human cancer cells lacking DNMT1. *Nature* *404*, 1003–1007.

Rhee, I., Bachman, K.E., Park, B.H., Jair, K.-W., Yen, R.-W.C., Schuebel, K.E., Cui, H., Feinberg, A.P., Lengauer, C., Kinzler, K.W., et al. (2002). DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature* 416, 552–556.

Rieder, L.E., and Reenan, R. a (2012). The intricate relationship between RNA structure, editing, and splicing. *Semin. Cell Dev. Biol.* 23, 281–288.

Robertson, K.D., Uzvolgyi, E., Liang, G., Talmadge, C., Sumegi, J., Gonzales, F.A., and Jones, P.A. (1999). The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors. *Nucleic Acids Res.* 27, 2291–2298.

Robertson, K.D., Ait-Si-Ali, S., Yokochi, T., Wade, P. a, Jones, P.L., and Wolffe, a P. (2000). DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. *Nat. Genet.* 25, 338–342.

Roscigno, R.F., and Garcia-Blanco, M. a (1995). SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. *RNA* 1, 692–706.

Rosonina, E., and Blencowe, B.J. (2004). Analysis of the requirement for RNA polymerase II CTD heptapeptide repeats in pre-mRNA splicing and 3' -end cleavage Analysis of the requirement for RNA polymerase II CTD heptapeptide repeats in pre-mRNA splicing and 3' -end cleavage. 581–589.

Rosonina, E., Bakowski, M. a, McCracken, S., and Blencowe, B.J. (2003). Transcriptional activators control splicing and 3'-end cleavage levels. *J. Biol. Chem.* 278, 43034–43040.

Rothbart, S.B., Krajewski, K., Nady, N., Tempel, W., Xue, S., Badeaux, A.I., Barsyte-Lovejoy, D., Martinez, J.Y., Bedford, M.T., Fuchs, S.M., et al. (2012). Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. *Nat. Struct. Mol. Biol.* 19, 1155–1160.

Rueter, S.M., Dawson, T.R., and Emeson, R.B. (1999). Regulation of alternative splicing by RNA editing. *Nature* 399, 75–80.

Saha, A., Wittmeyer, J., and Cairns, B.R. (2006). Chromatin remodelling: the industrial revolution of DNA around histones. *Nat. Rev. Mol. Cell Biol.* 7, 437–447.

Saint-André, V., Batsché, E., Rachez, C., and Muchardt, C. (2011). Histone H3 lysine 9 trimethylation and HP1 γ favor inclusion of alternative exons. *Nat. Struct. Mol. Biol.* 18, 337–344.

Saito, M., and Ishikawa, F. (2002). The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. *J. Biol. Chem.* 277, 35434–35439.

Saldanha, R., Mohr, G., Belfort, M., and Lambowitz, A.M. (1993). Group I and group II introns. *FASEB J.* 7, 15–24.

Saltzman, A.L., Pan, Q., and Blencowe, B.J. (2011). Regulation of alternative splicing by the core spliceosomal machinery. *Genes Dev.* 25, 373–384.

Sasai, N., Matsuda, E., Sarashina, E., Ishida, Y., and Kawaichi, M. (2005). Identification of a novel BTB-zinc finger transcriptional repressor, CIBZ, that interacts with CtBP corepressor. *Genes Cells* 10, 871–885.

Sasai, N., Nakao, M., and Defossez, P.-A. (2010). Sequence-specific recognition of methylated DNA by human zinc-finger proteins. *Nucleic Acids Res.* 38, 5015–5022.

Saxonov, S., Berg, P., and Brutlag, D.L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. U. S. A.* 103, 1412–1417.

Schafer, M.P., Boyd, C.D., Tolstoshev, P., and Crystal, R.G. (1980). Structural organization of a 17 kb segment of the alpha2 collagen gene: evaluation by R loop mapping. *Nucleic Acids Res.* 2241–2253.

Schor, I.E., Rascovan, N., Pelisch, F., Alló, M., and Kornblihtt, A.R. (2009). Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. *Proc. Natl. Acad. Sci. U. S. A.* 106, 4325–4330.

Schwartz, S., Meshorer, E., and Ast, G. (2009). Chromatin organization marks exon-intron structure. *Nat. Struct. Mol. Biol.* 16, 990–995.

Sewalt, R.G.A.B., Lachner, M., Vargas, M., Hamer, K.M., den Blaauwen, J.L., Hendrix, T., Melcher, M., Schweizer, D., Jenuwein, T., and Otte, A.P. (2002). Selective interactions between vertebrate polycomb homologs and the SUV39H1 histone lysine methyltransferase suggest that histone H3-K9 methylation contributes to chromosomal targeting of Polycomb group proteins. *Mol. Cell. Biol.* 22, 5539–5553.

Shah, Z.H., Ahmed, S.U., Ford, J.R., Allison, S.J., Knight, J.R.P., and Milner, J. (2012). A deacetylase-deficient SIRT1 variant opposes full-length SIRT1 in regulating tumor suppressor p53 and governs expression of cancer-related genes. *Mol. Cell. Biol.* 32, 704–716.

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T. a, Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., et al. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* 450, 908–912.

Sharma, P., Azebi, S., England, P., Christensen, T., Møller-Larsen, A., Petersen, T., Batsché, E., and Muchardt, C. (2012). Citrullination of histone H3 interferes with HP1-mediated transcriptional repression. *PLoS Genet.* 8, e1002934.

Sharma, S., Kohlstaedt, L. a, Damianov, A., Rio, D.C., and Black, D.L. (2008). Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. *Nat. Struct. Mol. Biol.* 15, 183–191.

Sharp, P.A. (1994). Split Genes and RNA Splicing Nobel Lecture. *Cell* 77, 805–815.

Shepard, P.J., and Hertel, K.J. (2009). Protein family review The SR protein family. 1–9.

Shepard, J., Reick, M., Olson, S., Brenton, R., and Graveley, B.R. (2002). Characterization of U2AF 6 , a Splicing Factor Related to U2AF 35 Characterization of U2AF 26 , a Splicing Factor Related. 22.

Shi, J., Zhang, T., Zhou, C., Chohan, M.O., Gu, X., Wegiel, J., Zhou, J., Hwang, Y.-W., Iqbal, K., Grundke-Iqbal, I., et al. (2008). Increased dosage of Dyrk1A alters alternative splicing factor (ASF)-regulated alternative splicing of tau in Down syndrome. *J. Biol. Chem.* 283, 28660–28669.

Shi, Y., Reddy, B., and Manley, J.L. (2006). PP1/PP2A phosphatases are required for the second step of Pre-mRNA splicing and target specific snRNP proteins. *Mol. Cell* 23, 819–829.

Shin, J.H., Li, R.W., Gao, Y., Baldwin, R., and Li, C. (2012). Genome-wide ChIP-seq mapping and analysis reveal butyrate-induced acetylation of H3K9 and H3K27 correlated with transcription activity in bovine cells. *Funct. Integr. Genomics* 12, 119–130.

Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* 311, 844–847.

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 479, 74–79.

Sims, R.J., Millhouse, S., Chen, C.-F., Lewis, B. a, Erdjument-Bromage, H., Tempst, P., Manley, J.L., and Reinberg, D. (2007). Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. *Mol. Cell* 28, 665–676.

Singh, J., and Padgett, R. a (2009). Rates of in situ transcription and splicing in large human genes. *Nat. Struct. Mol. Biol.* 16, 1128–1133.

Singh, R., Valcárcel, J., and Green, M.R. (1995). Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science* 268, 1173–1176.

Smallwood, A., Estève, P., Pradhan, S., and Carey, M. (2007). Functional cooperation between HP1 and DNMT1 mediates gene silencing. *Genes Dev.* 21, 1169–1178.

Smallwood, A., Hon, G.C., Jin, F., Henry, R.E., Espinosa, J.M., and Ren, B. (2012). CBX3 regulates efficient RNA processing genome-wide. *Genome Res.* 22, 1426–1436.

Smolle, M., and Workman, J.L. (2013). Transcription-associated histone modifications and cryptic transcription. *Biochim. Biophys. Acta* 1829, 84–97.

Spies, N., Nielsen, C.B., Padgett, R. a, and Burge, C.B. (2009). Biased chromatin signatures around polyadenylation sites and exons. *Mol. Cell* 36, 245–254.

Stevens, S.W., Ryan, D.E., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D., and Abelson, J. (2002). Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol. Cell* 9, 31–44.

Strahl, B.D., Briggs, S.D., Brame, C.J., Caldwell, J. a, Koh, S.S., Ma, H., Cook, R.G., Shabanowitz, J., Hunt, D.F., Stallcup, M.R., et al. (2001). Methylation of histone H4 at arginine 3 occurs in vivo and is mediated by the nuclear receptor coactivator PRMT1. *Curr. Biol.* 11, 996–1000.

Tachibana, M., Sugimoto, K., Fukushima, T., and Shinkai, Y. (2001). Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. *J. Biol. Chem.* 276, 25309–25317.

Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H., et al. (2002). G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev.* 16, 1779–1791.

Tachibana, M., Ueda, J., Fukuda, M., Takeda, N., Ohta, T., Iwanari, H., Sakihama, T., Kodama, T., Hamakubo, T., and Shinkai, Y. (2005). Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. *Genes Dev.* 19, 815–826.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W. a, Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930–935.

Tai, H., Geisterfer, M., Bell, J., Moniwa, M., Davie, J.R., Boucher, L., and McBurney, M.W. (2003). CHD1 associates with NCoR and histone deacetylase as well as with RNA splicing proteins. *Biochem. ...* 308, 170–176.

Talbert, P.B., and Henikoff, S. (2010). Histone variants--ancient wrap artists of the epigenome. *Nat. Rev. Mol. Cell Biol.* 11, 264–275.

Tan, M., Luo, H., Lee, S., Jin, F., Yang, J.S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., et al. (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* 146, 1016–1028.

Tardiff, D.F., Lacadie, S. a, and Rosbash, M. (2006). A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly posttranscriptional. *Mol. Cell* 24, 917–929.

Tennyson, C., Klamut, H., and Worton, R. (1995). The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. *Nat. Genet.* 9.

Tilghman, S.M., Tiemeier, D.C., Seidman, J.G., Peterlin, B.M., and Sullivan, M. (1978a). Intervening sequence of DNA identified in the structural portion of a mouse beta-globin gene. *Proc. Natl. Acad. Sci. U. S. A.* 75, 725–729.

Tilghman, S.M., Curtist, P.J., Tiemeier, D.C., Leder, P., and Weissmann, C. (1978b). The intervening sequence of a mouse B-globin gene is transcribed within the 15S B-globin mRNA precursor. *Proc. Natl. Acad. Sci. U. S. A.* 75, 1309–1313.

Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcárcel, J., and Guigó, R. (2009). Nucleosome positioning as a determinant of exon recognition. *Nat. Struct. Mol. Biol.* 16, 996–1001.

Tilgner, H., Knowles, D.G., Johnson, R., Davis, C. a, Chakraborty, S., Djebali, S., Curado, J., Snyder, M., Gingeras, T.R., and Guigó, R. (2012). Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* 22, 1616–1625.

Toffolo, E., Rusconi, F., Paganini, L., Tortorici, M., Pilotto, S., Heise, C., Verpelli, C., Tedeschi, G., Maffioli, E., Sala, C., et al. (2013). Phosphorylation of neuronal Lysine-Specific Demethylase 1LSD1/KDM1A impairs transcriptional repression by regulating interaction with CoREST and histone deacetylases HDAC1/2. *J. Neurochem.* n/a–n/a.

Tolstorukov, M.Y., Goldman, J. a, Gilbert, C., Ogryzko, V., Kingston, R.E., and Park, P.J. (2012). Histone variant H2A.Bbd is associated with active transcription and mRNA processing in human cells. *Mol. Cell* 47, 596–607.

Turunen, J.J., Niemelä, E.H., Verma, B., and Frilander, M.J. (2013). The significant other: splicing by the minor spliceosome. *Wiley Interdiscip. Rev. RNA* 4, 61–76.

Tyagi, A., Ryme, J., Brodin, D., Ostlund Farrants, A.K., and Visa, N. (2009). SWI/SNF associates with nascent pre-mRNPs and regulates alternative pre-mRNA processing. *PLoS Genet.* 5, e1000470.

Vakoc, C.R., Mandat, S. a, Olenchock, B. a, and Blobel, G. a (2005). Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. *Mol. Cell* 19, 381–391.

Vakoc, C.R., Sachdeva, M.M., Wang, H., and Blobel, G. a (2006). Profile of histone lysine methylation across transcribed mammalian chromatin. *Mol. Cell. Biol.* 26, 9185–9195.

Vargas, D.Y., Shah, K., Batish, M., Levandoski, M., Sinha, S., Marras, S. a E., Schedl, P., and Tyagi, S. (2011). Single-molecule imaging of transcriptionally coupled and uncoupled splicing. *Cell* 147, 1054–1065.

Velasco, G., Hubé, F., Rollin, J., Neuillet, D., Philippe, C., Bouzinba-Segard, H., Galvani, A., Viegas-Péquignot, E., and Francastel, C. (2010). Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. *Proc. Natl. Acad. Sci. U. S. A.* 107, 9281–9286.

Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The Spliceosome : Design Principles of a Dynamic RNP Machine. *Cell* 701–718.

Wahli, W., Dawid, I.B., Wyler, T., Weber, R., and Ryffel, G.U. (1980). Comparative analysis of the structural organization of two closely related vitellogenin in *X. laevis*. *Cell* 20, 107–117.

Waldholm, J., Wang, Z., Brodin, D., Tyagi, A., Yu, S., Theopold, U., Farrants, A.K.Ö., and Visa, N. (2011). SWI/SNF regulates the alternative processing of a specific subset of pre-mRNAs in *Drosophila melanogaster*. *BMC Mol. Biol.* 12, 46.

- Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476.
- Wang, L., Charoensuksai, P., Watson, N.J., Wang, X., Zhao, Z., Coriano, C.G., Kerr, L.R., and Xu, W. (2013). CARM1 automethylation is controlled at the level of alternative splicing. *Nucleic Acids Res.* 41, 6870–6880.
- Wang, Z., Zang, C., Cui, K., Schones, D.E., Barski, A., Peng, W., and Zhao, K. (2009). Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* 138, 1019–1031.
- Watt, F., and Molloy, P.L. (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev.* 2, 1136–1143.
- Webby, C., Wolf, A., Gromak, N., and Dreger, M. (2009). Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing. *Science* (80-.). 325, 90–93.
- Weinberg, M.S., Villeneuve, L.M., Ehsani, A.L.I., Amarzguioui, M., Aagaard, L., Chen, Z., Riggs, A.D., Rossi, J.J., and Morris, K. V (2006). The antisense strand of small interfering RNAs directs histone methylation and transcriptional gene silencing in human cells The antisense strand of small interfering RNAs directs histone methylation and transcriptional gene silencing in human cells. 256–262.
- Wiencke, J.K., Zheng, S., Morrison, Z., and Yeh, R.-F. (2007). Differentially expressed genes are marked by histone 3 lysine 9 trimethylation in human cancer cells. *Oncogene* 27, 2412–2421.
- Will, C.L., and Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* 3, 1–23.
- Wu, H., and Zhang, Y. (2014). Reversing DNA Methylation: Mechanisms, Genomics, and Biological Functions. *Cell* 156, 45–68.
- Wu, H., Min, J., Lunin, V. V., Antoshenko, T., Dombrowski, L., Zeng, H., Allali-Hassani, A., Campagna-Slater, V., Vedadi, M., Arrowsmith, C.H., et al. (2010). Structural Biology of Human H3K9 Methyltransferases. *PLoS One* 5, e8570.
- Wysocka, J., Swigut, T., Xiao, H., Milne, T. a, Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P., et al. (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* 442, 86–90.
- Xiao, R., Tang, P., Yang, B., Huang, J., Zhou, Y., Shao, C., Li, H., Sun, H., Zhang, Y., and Fu, X.-D. (2012). Nuclear matrix factor hnRNP U/SAF-A exerts a global control of alternative splicing by regulating U2 snRNP maturation. *Mol. Cell* 45, 656–668.
- Yang, W., Xia, Y., Hawke, D., Li, X., Liang, J., Xing, D., Aldape, K., Hunter, T., Alfred Yung, W.K., and Lu, Z. (2012). PKM2 Phosphorylates Histone H3 and Promotes Gene Transcription and Tumorigenesis. *Cell* 150, 685–696.
- Yoder, J. a, Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* 13, 335–340.
- Yoon, H.-G., Chan, D.W., Reynolds, A.B., Qin, J., and Wong, J. (2003). N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso. *Mol. Cell* 12, 723–734.
- Young, J.I., Hong, E.P., Castle, J.C., Crespo-Barreto, J., Bowman, A.B., Rose, M.F., Kang, D., Richman, R., Johnson, J.M., Berget, S.M., et al. (2005). Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. *Proc. Natl. Acad. Sci. U. S. A.* 102, 17551–17558.
- Yu, Y., Song, C., Zhang, Q., DiMaggio, P.A., Garcia, B.A., York, A., Carey, M.F., and Grunstein, M. (2012). Histone H3 lysine 56 methylation regulates DNA replication through its interaction with PCNA. *Mol. Cell*.

- Zeng, C., Kim, E., Warren, S.L., and Berget, S.M. (1997). Dynamic relocation of transcription and splicing factors dependent upon transcriptional activity. *EMBO J.* *16*, 1401–1412.
- Zhang, J., Gao, Q., Li, P., Liu, X., Jia, Y., Wu, W., Li, J., Dong, S., Koseki, H., and Wong, J. (2011). S phase-dependent interaction with DNMT1 dictates the role of UHRF1 but not UHRF2 in DNA methylation maintenance. *Cell Res.* *21*, 1723–1739.
- Zhang, Y., Ng, H., Erdjument-bromage, H., Tempst, P., Bird, A., and Reinberg, D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev.* *13*, 1924–1935.
- Zhang, Z., Jones, A., Joo, H.-Y., Zhou, D., Cao, Y., Chen, S., Erdjument-Bromage, H., Renfrow, M., He, H., Tempst, P., et al. (2013). USP49 deubiquitinates histone H2B and regulates cotranscriptional pre-mRNA splicing. *Genes Dev.* *27*, 1581–1595.
- Zhou, H.-L., Hinman, M.N., Barron, V. a, Geng, C., Zhou, G., Luo, G., Siegel, R.E., and Lou, H. (2011). Hu proteins regulate alternative splicing by inducing localized histone hyperacetylation in an RNA-dependent manner. *Proc. Natl. Acad. Sci. U. S. A.* *108*, E627–35.
- Zhu, J., and Krainer, A. (2000). Pre-mRNA splicing in the absence of an SR protein RS domain. *Genes Dev.* *14*, 3166–3178.
- Zibetti, C., Adamo, A., Binda, C., Forneris, F., Toffolo, E., Verpelli, C., Ginelli, E., Mattevi, A., Sala, C., and Battaglioli, E. (2010). Alternative splicing of the histone demethylase LSD1/KDM1 contributes to the modulation of neurite morphogenesis in the mammalian nervous system. *J. Neurosci.* *30*, 2521–2532.
- Zippo, A., Serafini, R., Rocchigiani, M., Pennacchini, S., Krepelova, A., and Oliviero, S. (2009). Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation. *Cell* *138*, 1122–1136.